世纪前沿

[美] 赫伯特·金迪斯　萨缪·鲍尔斯等 著

跨学科社会科学研究论丛

汪丁丁　叶 航　罗卫东 主编

人类的趋社会性及其研究

一个超越经济学的经济分析

上海世纪出版集团

世纪出版

ISBN 7-208-05776-1
9 787208 057760 >
定价：24.00元
易文网：www.ewen.cc

人类的趋社会性及其研究

一个超越经济学的经济分析

[美] 赫伯特·金迪斯　萨缪·鲍尔斯等 著

浙江大学跨学科社会科学研究中心 译

世纪出版集团　上海人民出版社

图书在版编目(CIP)数据

人类的趋社会性及其研究：一个超越经济学的经济分析/(美)金迪斯 (Gintis, H.)等著；浙江大学跨学科社会科学研究中心译.—上海：上海人民出版社，2005

ISBN 7-208-05776-1

Ⅰ. 人… Ⅱ. ①金…②浙… Ⅲ. 人的社会性—文集 Ⅳ. B038-53

中国版本图书馆 CIP 数据核字(2005)第 081447 号

出 品 人　施宏俊
责任编辑　王志毅
装帧设计　陆智昌

人类的趋社会性及其研究：一个超越经济学的经济分析
［美］金迪斯　鲍尔斯　等著
浙江大学跨学科社会科学研究中心　译

出　　版　世纪出版集团　上海人民出版社
　　　　　(200001　上海福建中路 193 号　www.ewen.cc)
出　　品　世纪出版集团　北京世纪文景文化传播有限公司
　　　　　(100027 北京朝阳区幸福一村甲 55 号 4 层)
发　　行　世纪出版集团发行中心
印　　刷　山东新华印刷厂临沂厂
开　　本　635×965 毫米　1/16
印　　张　14.75
插　　页　4
字　　数　180,000
版　　次　2006 年 5 月第 1 版
印　　次　2006 年 5 月第 1 次印刷
ISBN 7-208-05776-1/C·214
定　　价　24.00 元

出版说明

自中西文明发生碰撞以来，百余年的中国现代文化建设即无可避免地担负起双重使命。梳理和探究西方文明的根源及脉络，已成为我们理解并提升自身要义的借镜，整理和传承中国文明的传统，更是我们实现并弘扬自身价值的根本。此二者的交汇，乃是塑造现代中国之精神品格的必由进路。世纪出版集团倾力编辑世纪人文系列丛书之宗旨亦在于此。

世纪人文系列丛书包涵“世纪文库”、“世纪前沿”、“袖珍经典”、“大学经典”及“开放人文”五个界面，各成系列，相得益彰。

“厘清西方思想脉络，更新中国学术传统”，为“世纪文库”之编辑指针。文库分为中西两大书系。中学书系由清末民初开始，全面整理中国近现代以来的学术著作，以期为今人反思现代中国的社会和精神处境铺建思考的进阶；西学书系旨在从西方文明的整体进程出发，系统译介自古希腊罗马以降的经典文献，借此展现西方思想传统的生发流变过程，从而为我们返回现代中国之核心问题奠定坚实的文本基础。与之呼应，“世纪前沿”着重关注二战以来全球范围内学术思想的重要论题与最新进展，展示各学科领域的新近成果和当代文化思潮演化的各种向度。“袖珍经典”则以相对简约的形式，收录名家大师们在体裁和风格上独具特色的经典作品，阐幽发微，意趣兼得。

遵循现代人文教育和公民教育的理念，秉承“通达民情，化育人心”的中国传统教育精神，“大学经典”依据中西文明传统的知识谱系及其价值内涵，将人类历史上具有人文内涵的经典作品编辑成为大学教育的基础读本，应时代所需，顺时势所趋，为塑造现代中国人的人文素养、公民意识和国家精神倾力尽心。“开放人文”旨在提供全景式的人文阅读平台，从文学、历史、艺术、科学等多个面向调动读者的阅读愉悦，寓学于乐，寓乐于心，为广大读者陶冶心性，培植情操。

“大学之道，在明明德，在新民，在止于至善”（《大学》）。温古知今，止于至善，是人类得以理解生命价值的人文情怀，亦是文明得以传承和发展的精神契机。欲实现中华民族的伟大复兴，必先培育中华民族的文化精神；由此，我们深知现代中国出版人的职责所在，以我之不懈努力，做一代又一代中国人的文化脊梁。

上海世纪出版集团

世纪人文系列丛书编辑委员会

2005年1月

目录

何谓“社会科学根本问题”？
——为“跨学科社会科学研究论丛”序/1　汪丁丁

导读一：人类合作秩序的起源与演化/8　汪丁丁　罗卫东　叶　航

导读二：作为内生偏好的利他行为及其经济学意义/27
叶　航　汪丁丁　罗卫东

作者中译本序/48

人类合作的起源/52

社会资本和共同体治理/69

共同体的道德经济：结构化的人群和趋社会规范的演化/92

私有财产的演化/117

不平等的遗传/142

人类利他行为的解释/177

附录：利他惩罚的神经基础/202

译名对照表/216

后记/220

何谓“社会科学根本问题”？
——为“跨学科社会科学研究论丛”序

汪丁丁

为什么要追问“根本问题”？借用黑格尔的语言，一个核心概念在各向度上的充分展开，就是全部理论。再借用西美尔的类比，一旦康德开始追问“自然如何可能”的时候，他就写出了《纯粹理性批判》；当西美尔追问“社会如何可能”的时候，他写出了一系列“社会理论”的基础论文；当我们追问社会科学根本问题的时候，我们就可能写出统一的社会科学。

在上述视角下审查人类知识的进步，我们意识到，全部人类知识，其实就是以追问根本问题的方式被激发、获取和积累起来的。人类知识的起点——哲学，始于对未知的“敬畏”，始于“爱智”，始于“天问”。从哲学当中，对“天”与“人”之间关系的追问，导致了道德哲学；由此又发展出两种叙事方式，其一是“科学”——陈述外在感受，其二是“人文”——陈述内在感受。叙事方式是思维的惯式——思维可以是一，惯式却可以有多。

沿着科学叙事的传统，对宇宙起源问题的关注导致了古代希腊语词所谓的“物理学”——在这一学科内部，根本问题分殊为各层次的和各方面的，就产生了星象学、几何学、化学、数学。对人类起源问题的关注导致了古代希腊语词所谓的“生物学”——在这一学科内部，根本

问题分殊为各层次的和各方面的，就产生了博物学、遗传学、医学。

沿着人文叙事的传统，对灵魂起源问题的关注导致了古代希腊语词所谓的“心理学”——在这一学科内部，根本问题分殊为各层次的和各方面的，就产生了丧葬、图腾、神话、神学、命理学。对思维与叙事方式的关注导致了古代希腊语词所谓的“逻辑学”（逻各斯）——在这一学科内部，根本问题分殊为各层次的和各方面的，就产生了修辞学、名学、语言学、符号学、脑科学。

第三，也是我们目前最为关注的，在科学叙事与人文叙事这两种思维惯式之间，始终存在着所谓的“跨学科”思维惯式。在西方思想史上，所谓“社会科学”，肇端于19世纪中叶，在生物学思想的主导下发展成为今天我们见到的样式。今天，社会科学领域内最偏向于科学叙事的部分，可以概括在“行为学”这一名称之下——它把人降低到动物现象的层次上加以研究。另一方面，社会科学领域内的最偏向于人文叙事的部分，可以概括在“伦理学”这一名称之下——它把人提高到精神现象的层次上加以研究。

科学的与人文的叙事传统，以及传统内部积累起来的人类知识，依照康德的分类，呈现出沿时间的秩序和沿空间的秩序。前者称为“历史”，后者称为“结构”。

把历史与结构应用于社会科学，所谓“根本问题”，就表现为贯穿着社会科学全部历史的结构问题。我们不打算论证，也不太相信社会科学的根本问题是惟一的。有鉴于此，我们采取了枚举法，来论证社会科学根本问题的存在性。

首先，社会科学是建立在关于“社会”的经验基础上的知识，故而在它的传统之内，它只承认获得了经验支持的知识表达。每一位认真的社会科学研究者都难以拒绝这样一项陈述：贯穿着人类社会以及社会性动物社会的历史的一类秩序——通常被称为“合作”——不论从行为学角度审视还是从伦理学角度审视，它对社会现象而言，都具有“根本”的意义。

其次，社会科学的知识，与科学知识类似，必须表达为关于“结构”的陈述。对于“合作”这类社会现象，每一位认真的社会科学研究者都难以拒绝关于合作秩序的结构的知识。因为合作的秩序，尤其是它的空间形态，是一切社会现象在科学叙事传统内获得令人信服的解释的基础。

这样，我们不妨从一个特定角度把社会科学根本问题定义为：“何种结构导致了合作？”注意，这里提出的定义，仅仅是从上述的特定角度提出来的。根据“对话的逻各斯”的逻辑，任何根本问题的定义都不是惟一的。

何种结构导致了合作？这一问句包含了两个关键概念，其一是“结构”——人类知识沿空间的秩序，其二是“合作”。后者需要进一步界定，以免“社会科学”泛滥为“科学”。

古希腊人的科学叙事，据海德格尔考证，其特征在于把事物的“本质”放置于事物的发生演变过程当中，从而避免落入后来西方人落入的思维的形而上学陷阱。把合作秩序放置于演化过程中，这样获得的知识，我称之为“合作的发生学”。

就语义而言，“合作”指称的，是个体与个体之间的一种关系，这种关系在经验世界里与“竞争”关系相对比构成了足够显著的差异，以致我们更愿意把它命名为“合作”而不命名为“竞争”。2004 年 10 月发表在《神经呈像》杂志上的一篇研究报告显示，合作关系和竞争关系激活了不同的脑区组合。这一事实表明，经过漫长的物种演化和社会演化，今天，合作与竞争的神经元网络很可能激发出具有本质差异的人类情感。

合作现象的发生学要求我们在界定“合作”之前首先界定“个体”。如果我们关注的个体是人类个体，那么，合作就应当被理解为是人与人之间关系的某类结构所导致的社会现象。如果我们关注的个体是单细胞，那么，合作就应当被理解为是细胞与细胞之间关系的某类结构所导致的社会现象。例如，生物学家非常熟悉的“共生”现象，

似乎是一种合作。当然，共生现象也可以被看作是一种由于相持不下而实现的竞争的均衡。

其实，共生现象为我们提供了一个出色的例子，来说明竞争与合作的结构辩证法和广义政治学。当竞争着的个体达成某种合作秩序时，从更高层次观察，它们似乎构成了一个更大的个体——“社会的个体”。另一方面，走进任何一个社会，我们都可以观察到社会成员之间的竞争关系。我们甚至愿意承认：正是个体之间的竞争关系，界定了社会内部的“个体”概念。

于是，当竞争关系对观察者而言不十分显著时，“个体”就消失，取而代之的，是“群体”。例如，典型意义上的“植物”——相对“动物”而言，被定义为“缺乏个体性的物种”。

这样，基于上述“个体”与“群体”的关系辩证法，我们甚至可以声称：所谓“合作的结构”，无非是与“竞争的结构”相比较而言，凸显为合作的那些结构。例如，在一家企业内部，其实存在着个体之间的激烈竞争，但就与市场里相互竞争着的许多企业之间的关系而言，企业内部的这些竞争关系可以被忽略，代之以合作关系。

更严格地说，一位经验主义者，或许可以把“社会”定义为：在单位时间内被观察到的不同时空点处的事件之间发生的足以引发观察者因果性联想的关系的频率达到了被观察者认为显著的程度，这些事件的全体，就构成一个社会。

不过，西美尔非常不赞同从观察者角度来定义“社会”，因为那样就有可能忽略社会成员之间的心理联系。作为对西美尔的批评的回应，我们可以把上面给出的经验主义的“社会”定义稍加拓展，让它能够包含“心理联系”——不同时空点处的事件之间的因果关系不仅是物理的而且可以是心理的。纯粹的心理联系，涵盖着西美尔定义过的“可社会性”。

社会如何可能？这一西美尔问题导致了“广义社会理论”的发展。今天，我们或许可以这样回答西美尔问题：社会因个体之间的合

作关系而成为可能。注意，这一回答导致了"社会科学"，从而比西美尔的理论更狭义和更具实证性。同时，这一回答所包含的说服力超过了另一种似乎与它等价的回答——"社会因个体之间的竞争关系而成为可能"。

当代经济学，由于它与"个体理性"概念和"理性选择"概念之间的密切关系而在社会科学诸学科当中占有一种特殊的位置。从古典政治学家如斯密和小密尔，到新古典经济学家如萨缪尔森和贝克尔，再到公共选择理论家如阿罗、森、布坎南，我们看到一群出类拔萃的经济学家，他们不仅关注经济学问题而且关注社会科学的根本问题。

在经济学文献中，合作问题通常由"囚徒困境"一次博弈来刻画。近年来，一些学者开始研究这一博弈的连续策略解，并在"合作"与"不合作"这两个极端之间引入连续的"合作度"。这样，合作就可以被视为基于生存竞争的个体理性选择，随着所选策略连续地趋于合作解，竞争着的一群个体就逐渐构成一个协调着的整体——"社会的个体"。

纳什最早把合作博弈刻画为竞争着的个体的理性选择过程——所谓"二人讨价还价问题"的解，及稍后发表的定义在"威胁"的策略集上的"二人合作博弈问题"的解。

在包括经济学在内的 1 661 种 1995 年以来以英文出版的科学与社会科学期刊的大约 13.4 万期文章中搜索关键词"合作"，我得到了 4 209 篇学术论文。以显著频率出现在这批文献里，与"囚徒困境博弈"的合作解密切相关的一个概念，是所谓"对等性"——英文是"reciprocity"，依场合不同常被译作"交互性"、"互惠性"、"己所不欲，勿施于人"。

例如，在国际贸易与国际关系的研究中，一份 2004 年 9 月发表的报告表明，国家之间的互惠交往是国际间合作关系的最重要因素之一。如果两国间地理距离的增加使交易费用上升，那么，两国间合作关系弱化的速度将比两国间冲突关系弱化的速度更慢些。换句话说，长期而

言，国际关系呈现出来的格局不是“远交近攻”而是“远攻近交”。推广而言，对地球人来说，外星人与地球人之间发生冲突的概率将大于地球上各国之间发生冲突的概率。用社会学的语言解释这件事情，就是语言和交往导致了更多的了解。后者为人类的合作关系提供了情感的与理性的基础，并且因此而发生的合作效应比因交往而发生的冲突效应更强烈。基于同样的原则，一个组织，它内部的人际之间的合作效应肯定比冲突效应更强烈。也因此，我们才可能观察到这一“组织”。

互惠性，在一位晚近获得诺贝尔奖的经济学家的解释中，更主要的是一种相互惩罚的可能性——如果一方背叛了合作，那么另一方就有惩罚背叛者的冲动和权利。因为，对于方法论个人主义者而言，基于单纯互惠性的人类合作几乎是不可思议的。金迪斯等人的研究表明，在关于合作的“囚徒困境”博弈的策略集合内引入“惩罚”策略，可以极大地扩展合作的空间尺度和时间尺度。

在一篇发表于 2004 年 11 月的论文中，接续着金迪斯等人的思路，作者指出，对背叛合作者的惩罚，在组织内部比在组织外部更强烈。也因此，在地球上，同一区域内相邻各国之间的冲突关系甚至比它们之间的合作关系更为显著——因惩罚变得更加残酷而更显著。

翻译和收录在这套论丛的第一册（《走向统一的社会科学》）里的六篇论文，它们的主要作者是金迪斯教授和鲍尔斯教授——两位长期合作研究的作者。金迪斯 1969 年在哈佛大学获得经济学博士学位，1962 年在哈佛大学获得数学硕士学位，1961 年在宾州大学获得数学学士学位。鲍尔斯 1965 年在哈佛大学获得经济学博士学位，1960 年在耶鲁大学获得文科学士学位。鲍尔斯教授曾发表过论文批评萨缪尔森，并因此为我所知。此外，他 1998 年发表于权威刊物《经济学文献杂志》（*Journal of Economic Literature*）上的综述型论文“内生偏好”，引起学界的广泛注意。自 2000 年起，鲍尔斯教授转任桑塔费研究院“经济学板块”的研究主任，从而成为所谓“桑塔费经济学”的代表人物。金迪斯在麻省大学任教多年，晚近研究演化博弈论和博弈

论的演化并发表了专著《博弈论的演化》（*Game Theory Evolving*）。荣休之后，他转至桑塔费研究院，已经在那里发表了多篇工作论文。2005年1月，作为第一主编，他与鲍尔斯等三位主编共同出版了文集《道德情操与物质利益》（MIT出版社）。这部文集意味着西方学术界比以往更加关注斯密《道德情操论》的思想遗产，并试图在包括脑科学在内的最新研究成果的基础上，对长期以来被主流经济学过分关注和扭曲了的斯密《国富论》的思想遗产加以反省。

收录在这里的论文，旨在解释人类社会的一种超越普通动物界的现象——广泛存见于人类社会而不见于非人类社会的非亲缘个体之间的合作关系。作者们的这一努力，持续约十年时间，在至少五门不同的自然科学与社会科学之间进行合作研究，今天，借助于脑科学研究手段，已经有了突破性的进展。关于这一突破性进展，在这套论丛的第二册（《人类的趋社会性及其研究》）里收录了更细致的研究报告。

跨学科的努力，为着解答社会科学的一个根本问题——何种结构导致了合作？这是一个跨学科的问题，故而要求跨学科的合作研究。今天，根据我们的文献阅读，西方学者们正在从经济学、生物学、社会学、文化人类学、演化心理学、认知科学和符号逻辑等领域，围绕合作的发生学问题进行合作研究。

通过这套论丛，我们由衷地希望国内读者注意到社会科学研究在过去十年所经历的这一方向性转变，注意到这一转变很可能引发的革命性后果，并尽快参与这些预期将成为社会科学前沿课题的研究。还是那句话：我们把这套论丛献给未来的社会科学家们。

导读一：

人类合作秩序的起源与演化*

汪丁丁　罗卫东　叶　航

一

汪丁丁：我想从两个角度，把这次报告的主题引出来。其实，我们今天的报告是一个跨学科问题。阐释这个问题，仅仅依靠经济学是不够的。我们必须把它放在整个社会科学的视角上，用一种统一和演化的社会理论来谈人类的合作秩序。

从演化社会理论谈这个问题，有两个角度可以切入。第一，我们假设，一个外星人突然来到我们地球，他要做的第一件事当然是想知道，统治这个星球的动物，即我们人类究竟是怎么回事。作为外星人，他不会去考虑你是怎么行动的，你的动机是什么，或者你的行为背后是什么理论体系。他要把握的其实是最简单、最直观或者最宏观的事实，即这种动物以什么方式来实现他们之间的联系？从这个角度，外星人也许会发现，人类首先是以个体为单位进行活动的，不像其他动

* 本文是三位作者根据他们 2005 年 4 月 22—24 日在南京理工大学的演讲内容修改整理而成，发表于《社会科学战线》2005 年第三期。

物那样一群一群地居住。维系人类个体之间联系的有两种主要形态，第一种叫竞争关系，这是经济学家很强调的关系，它导致了效率，导致了很多很多或好或坏的东西。但外星人显然还会注意到另外一种关系，即人类之间的合作关系。就我这个地球人来说，我不知道人类的这两种关系究竟哪个更重要。我到现在还没有想清楚，从发生学的角度看，竞争关系和合作关系究竟哪一个先，哪一个后。但我可以找出最新的研究文献，表明人类对合作的兴趣，最起码在生物脑演化的阶段上早于人类的竞争关系。你别看我们经济学家谈竞争谈了那么多，其实合作的重要性不亚于竞争的重要性，只不过经济学家不谈，或者很少谈到。这是主流经济学很大的一个不足，是需要我们来改变的。

我们今天的主题是讨论合作关系，因为竞争关系谈得太多了。就竞争关系而言，在芝加哥学派的价格理论里面经常讲到，经济学的原理只有一个核心的概念——当然这是就芝加哥学派的口述传统而言的，一般教科书很少把它当成一个核心的概念——那就是“替代性”。各位在教科书里面看到“替代性”的时候可能会觉得很平常、很普通，其实它是主流经济学的一个核心概念。所谓价格理论、消费理论、生产理论、机会成本甚至理性选择等等，其实都可以追溯到资源之间或者事物之间的可替代性上。但我们今天谈的是合作关系，那么合作关系的核心概念是什么？相对于竞争关系的核心概念，我们有一个关于合作关系的核心概念，即“互补性”。这个在芝加哥学派的口述传统里面也出现很多次了，只不过像加里·贝克尔这样的领袖人物注意到了，但是他不怎么说，也不怎么写文章谈这件事情。就人类的合作关系来说，“互补性”是非常重要的。因为万事万物之间，不仅仅有非此即彼的“替代性”，而且还有相辅相成的“互补性”。前者导致了我们人类的竞争关系，而后者导致了我们人类的合作关系。

这两个核心概念我把它放在这儿了，它们的含义是什么？从个体之间的竞争关系我们可以推出的核心概念是“替代性”，它意味着个性的发展；事物的可替代性导致竞争，竞争导致专业化，专业化导致今天

气象万千的个性化世界。从个体之间的合作关系我们可以推出的核心概念是"互补性"，它意味着群体的发展；事物的可互补性导致合作，合作导致社会化，社会化导致今天气象万千的共生化世界。

上面是我说的第一个角度，再从第二个角度看我们今天报告的主题。外星人走了，现在轮到我们地球人自己解释这件事情了。于是就有所谓的理论家出来发表言论，理论家的任务是要从现象和经验观察中找出规律性的东西，比如通过观察天体运动推出牛顿三大定律，即给经验的世界建模。那么，传统的经济学家是怎么给人类行为建模的呢？他们建立的模型是所谓"理性选择"模型。这是大家非常熟悉的，我就不解释了，这是整个现代经济学的核心范式。

但现在的问题是，仅仅用所谓的"理性选择"是否能够解释人类的全部行为，包括我们刚才提到的那些在外星人看来显而易见的人类合作行为？社会学家和一部分重要的经济学家，甚至像马歇尔这样的现代经济学创始人都承认，有两个主要因素影响或者决定了人类的行为。哪两个因素？在中文传统里叫"情"与"理"。马歇尔在《经济学原理》第八版的前言里说，决定我们人类行为最根本、最长远的力量，一个是经济，另一个是宗教。[1] 什么是宗教？宗教就是一种情感。所以"情"与"理"是无法割裂的，在人类的选择行为中是两个基本的支点。我们人类的理性从来就不是冷酷的、不带情感的理性。

在以往的经济学教科书和主流经济学家的文章里，理性变成了不可爱的理性，变成了没有情感、社会正义和道德意识的理性，变成了冷酷的市场计算——"威尼斯商人身上的一磅肉"。我们浙江大学跨学科社会科学研究中心注意到了最近10年西方经济学理论发生的一种变化。这种变化很微妙，是由主流经济学家自己意识到的，并且由一些主流经济学家带头调整了方向，开始转到对人类情感的研究上。在这一研究方向上，除了经济学，还有脑科学、认知科学、文化人类学、社会心理学和演化心理学的参与。如果纯粹从经济学的角度看，这些经济学家正在试图把人类的情感因素综合到博弈论的框架里面来。

2001年，一份很重要的学术刊物《经济行为与组织》（*Journal of Economic Behavior and Organization*）上发表了一篇文章，题目叫“带有同情心的囚徒困境博弈”[2]，它的主要结论是同情心的存在可以在单次囚徒困境中导致合作。根据作者所做的博弈实验，在单次囚徒困境条件下，参与者的同情心越强，参与者之间同情共感的距离越近，合作就越容易实现。作者引进了一个度量心理距离的参数，结果发现这个参数与合作的概率完全成反比。人们的心理距离越小，合作发生的概率就越大。当这个参数收敛到某一个域值或者某一个点的时候，参与者几乎百分之百的合作。在这个点之前有一个区域，两个人之间的关系是一部分合作，一部分不合作。作者挑选的实验者，都是同一所大学的学生。挑选的标准是，他们之间必须是互相熟悉并共同相处一年以上，只有这样才可以度量他们的同情心。通过博弈实验，他们得到了一些非常可信的数据。比如，作者发现同情心并不是对称的，我同情你的时候，你未必就同等程度地同情我。玩这个游戏的时候出现了一种情况，作者把它叫做“同情者的礼物”：我情愿单方面和你合作，甚至明明知道你会背叛我、出卖我，我也毫无怨言地作出“牺牲”，仅仅因为我可怜你、爱你或者崇拜你。

上述现象都是把同情心引进了博弈论以后出现的新观察，或者是有待进一步研究的领域。这是个很有意思的事情，它说明我们的决策行为，甚至在单次囚徒困境条件下也是无法离开情感因素的。它说明，“情”这个东西，在今天主流经济学家的眼睛里，已经不再像以前西方传统里的“情”一样，是可以和“理”完全分开的。我们可以把这种研究称作“理”的情感化研究，它几乎和中国的传统看法一样，因为在中国人眼里，“情理”两个字从来就是合一的、连在一起的。

我们综述了很多西方的文献，加上中国人的理解。因为中国人和西方人处于两个极不相同的文化传统，很天然地成为相互的他者，或者是对方的镜子。在西方长大的西方经济学家、社会理论家，不能自觉地知道他自己的局限性。我们作为西方传统之外的学者，往往很容易看到他们的局限性，这样就可以加以补充，作出我们中国人自己的贡

献。另一方面，我们自己也受到自己文化传统的局限，不可能知道自己的弱点，无法跳出来，所以我们仍然要大量学习西方人的著作。

在"理"的情感化研究方向上，西方学者最近10年除了有许多非常前沿的探索之外，还有许多非常深刻的反思。他们开始反思，在经济学和社会理性选择理论的思想发展脉络上，他们究竟从什么时候开始走错了路。这是一件很重要的事情，是一种思想史的梳理。正是这种反思和疏理，把当代经济学家重新带回古典经济学家亚当·斯密的语境。当然，我说的不是亚当·斯密的《国富论》，而是亚当·斯密的《道德情操论》。西方经济学家发现，在过去200多年中，人们对斯密有太多误读，甚至完全背离了斯密最初对人类经济行为的洞见。所以西方学者提出，要重新发现斯密。最近10年来，西方人开始把亚当·斯密的研究变成一个非常热门的话题，有大量讨论亚当·斯密的文献出现。这是罗卫东老师博士论文的主题，等下罗老师会给我们详细介绍这方面的情况。前些年，有一家出版社请我为《拯救亚当·斯密》写一篇书评，这篇书评我写得很长。为什么？就是因为有感于上述变化。

在"理"的情感化研究方向上，还有另外一个方面的研究是完全承接西方人自己的套路的。西方的思想传统是逻各斯中心主义的，逻各斯在形式逻辑方面的表现就是数学，在技术手段方面的表现就是科学。这两个方面，基本上可以说就是"赛先生"。在"赛先生"教导的方向上，西方人仍然在继续往前走，即把人的选择行为当作一般动物的行为学模式来研究，因为一般动物是有情感冲动的。于是在这个方向上，很多动物学家、行为学家、行为心理学家、脑科学家都加入到这个研究行列里面来，用科学仪器研究人脑里"情"与"理"发生的科学过程。这是叶航老师关注的领域，等下叶老师会给我们介绍这方面的进展。

从2000年到现在，西方学者终于发现了"情"和"理"本来就是一回事，本来就是互相纠缠的，根本不可能像萨缪尔森所说的那样把它们一刀切开：这边是完全的理性选择，它解决的是最大化问题；那边是完全的情感冲动，它决定的是社会福利函数；理性选择给社会福利函数的

各个组成部分赋予不同的权重，然后把它们加起来。社会福利函数没有这种事情，今天的科学家不承认这种事情。这就是今天西方经济学和社会选择理论最新、最前沿的研究方向。在这个研究方向上，我们看到了两个世界：一个是思想史的世界，即亚当·斯密的世界；另一个是科学前沿的世界，即脑科学的世界。好了，我的引言就到这里。下面我把话筒交给罗老师和叶老师，让他们谈一下他们各自探索的世界。

二

罗卫东：丁丁为我们今天的论题开出了一个很合适的入口，这样就很顺利地引出了我们后面的话题。现在经济学正在酝酿一场革命。浙江大学跨学科社会科学研究中心的教授们一直在研究现代经济学范式所遇到的危机以及可能的出路，国内也有很多同行对此给予关注。在座的韦森教授身体力行也在做类似的工作。我个人在这个过程当中承担的角色主要偏重思想史方面，具体说是对现代经济学基本命题的古典思想资源进行追溯，尤其关注在现代经济学乃至整个社会科学中，有哪些资源可以为现代的经济学范式革新提供支持。

我不得不回到亚当·斯密，因为他是经济学的鼻祖。现代这么发达的经济科学，几乎全部的重要思想都来源于他的集大成著作，这是相当了不起的。亚当·斯密提出的很多命题都具有全新的、创造性的特点。斯密为人熟悉的是他对经济学的贡献，但我们必须知道，他本质上是一个综合性的社会科学家。他的全部思想和理论都是问题导向而非学科导向的。所以，斯密理论最关键的一个方面就是它的整体性，或者说它是没有学科思维定势的。在很长时间里，人们熟悉的是亚当·斯密的《国富论》，而对他的另外一部重要著作《道德情操论》不甚了了。其实，后者才是亚当·斯密本人生前更加重视的一本著作。遗憾的是，在他死后的100年里，人们几乎完全遗忘了《道德情操论》

这本书，经济学家为《国富论》着迷，对《道德情操论》毫无印象。但是，这个状况在上个世纪70年代以后开始发生巨大变化。

1976年，亚当·斯密的《国富论》（1776年）发表200周年纪念大会在英国格拉斯哥召开。格拉斯哥大学也是亚当·斯密生前担任过校长的大学，他还长期担任这个学校的道德哲学教授。以这次纪念活动为契机，有许许多多的学者开始超出《国富论》设定的范围，重新梳理斯密重要经济思想的哲学根源，由此开始关注他的《法学讲稿》、《道德情操论》等著作。借着这个纪念活动的推动，学术界推出了一个最权威的《亚当·斯密全集》，这个文本对最近30年亚当·斯密研究的推进提供了强大的支持。人们可以通过他没有面世的文稿和学生笔记来了解他的真实思想脉络。

亚当·斯密写《道德情操论》，花的时间和精力都是非常可观的。第一次出版是1759年，到最后一次修改是他死前三个月，1790年。中间有31年的时间，一共出了六个版本。他的《国富论》生前只出了三版。可以看出，他对《道德情操论》这个著作非常重视，始终不忘去完善它。实际上，伦理秩序是我们经济行为和政治制度相当重要的基础。如果我们对伦理方面不加研究，我们的经济行为和政治制度的"合法性"就很难说清楚。近代资本主义制度，就其主要的方面，特别是就其经济机制的设计而言，大体上是基于亚当·斯密《国富论》的体系。但是，市场机制带来效率的同时，也随之带来了道德情感方面的严重问题。对此，《国富论》并未作集中的探讨，倒是《道德情操论》讨论了这个问题。在斯密看来，商业社会道德感情的败坏是人性和制度合谋的结果，因此需要从两个方面加以关注。

我们全部的经济学基本假定，其核心是"经济人"假说。事实上，亚当·斯密从来没有提出过"经济人"假说，都是后人提炼出来的。厨师和面包师不是因为他的仁爱，而是他的自利才使我们每天能够吃到需要的食品，斯密的这句话常常被当作他主张自利是市场经济的人性基础的明证。在斯密看来，"看不见的手"，实际上是一个自利

的手，每个人按照自己利益的指引去生活，结果公共福利能够提高。斯密对自利人与市场关系的分析，是正确的，但是，也造成了误解。那时以来的经济学似乎不再问：为什么一个纯粹自利的人会选择交易这种和平与双赢的方式去对待他人？为什么两个自利的人彼此的行为一定会增进他人福利？难道这其中不需要某些必要的前提条件吗？我们看到，现在的经济学理论已经开始反思这个问题了。但是，经济学到现在才发问，这到底是因为斯密的误导还是我们对斯密的误解。我本人研究得出的结论表明，两个多世纪以来，我们的经济学多么严重地误解了斯密的本意，多么可怕地忽视了他思想中真正有价值的内容。

我要提醒大家注意的一点是：斯密《国富论》当中对自利人自由选择可以增进公共利益的命题，其真正的学理基础隐藏在他《道德情操论》关于同情心的重要思想之中。《国富论》和《道德情操论》并非互相独立的两部著作，通过仔细考察可以发现，《道德情操论》是斯密全部社会科学的方法论基础，《国富论》则是斯密将他的基本思想运用于财富研究所得出的成果。就学术重要性而言，《道德情操论》远在《国富论》之上，尽管后者在社会影响方面超过了前者。

让我们用最简单的囚徒困境模型来讨论《道德情操论》的意义。如果甲、乙两个囚徒是极端自私的行为者，那么必然会陷入不合作的纳什均衡，这已经是广为人知的结论了。然而，我们发现，在人类生活的各个领域，合作是一个常态。很多经济学家试图对这样一个悖论作出解释，比如“重复博弈”，或者调整博弈规则即改变报酬矩阵等等。但是在纯粹自利人的基本假设下，解决这个悖论是不可能的。这意味着，除非我们对人性的假设作出调整，否则便无法很好地解释人类的合作秩序如何发生这样一个问题。

我们之所以难以在自利人的假设基础上解释普遍的人类合作行为，一定是因为我们关于人的天性的假设遗漏了非常重要的东西。在古典经济学时代，这个东西还是在那里的，但是在边际革命以后，甚至早在西尼尔那里，这个东西就被略去了。在某种意义上，李嘉图时代就放

弃了关于丰富人性的假设，而采取了单一人性的假设。这个假设非常适合于将社会理论精致化，并赋予其科学的形式，所以被理性主义时代以来的科学家所热烈推崇。但是，关于自利人的片面假设被囚徒困境的纳什均衡拖入了沼泽地，其不合理的一面彻底地展示出来了。这就迫使今天的社会科学家尤其是经济学家思考那个被遗漏掉的人性理论。这样一来，我们就必须回到创立经济学之前的斯密，回到苏格兰启蒙学派的其他作者，如哈奇逊和休谟。

那么，斯密又是如何来看待这个问题的呢？人类社会之所以没有陷入囚徒困境，我们的合作之所以可以成功，一定是人类具有了某种超越自利的天性。在斯密看来，这种天性就是人的同情共感能力。所谓同情共感就是一个人对他人的喜怒哀乐有着即时的身心反应能力。我们能够对他人的遭遇感同身受，这个能力使得我们自然地形成某种心理和行为倾向，能够被他人的快乐和痛苦所感染。正是因为有这种能力，我们总是能够站在旁观者的立场上考虑和评判自己的行为，就何为合适的行为作出判断。

同情共感实际上就是人类形成社会秩序的基本的天然禀赋。也就是说，一个自利的人，同时也是一个具有同感能力的人，他在处世方面总是在自己的利益和他对别人感情的考虑之间作出判断和协调。这样一来，任何现实的个人并不只有自利的一面，而且还有设身处地为他人考虑的一面。如果人类仅仅具有自利的天性而没有同情同感的能力，就不一定会普遍选择交换作为获取利益的手段。因为纯粹自利人从效率的角度出发，更有可能选择暴力。斯密对人类行为中情感基础的考察为解释人类合作行为提供了新的理论基础。

翻看《道德情操论》，我们可以看到斯密在反复地探讨一个重要问题，那就是人类同情心的来源、形式和表现，以此为基础考察人类秩序的起源和运行。他翻来覆去说明的一个道理是：人类社会的秩序之所以可能，人类之所以能够合作，我们之所以能够组成社会，不仅因为我们自私，还因为我们时时刻刻都有某种设身处地为别人考虑的能力，始终都

有换位思考的天生禀赋，也就是我们一般而言的有同情心。这个研究虽然完全建基于观察和内省的经验，但在今天却是具有极为有益的启示。

斯密认为社会应该而且完全可能是建构在人类与生俱来的同情共感的天赋能力之上的。他从同情共感这个最基本的禀赋当中，发现了商业社会得以成立，且具有让人放心的性质的根据。斯密被人们看做苏格兰启蒙运动的代表性思想家之一，也是因为他从情感机制中找到了离开神的指导之后，人类能够生存和发展的理由。在他看来，人类凭着同情心就可以产生合作秩序，人与人之间可以通过同情心的互相作用，形成某种具有合宜性的规则和秩序。推而言之，市场经济的道德基础也是来自于人的天性当中的两层含义，即理性和情感。说到底，秩序或者规则是人们彼此之间的情感互动达到均衡的产物。

综上所述，对于今天经济学基础的重建来说，以亚当·斯密为代表的古典思想家的著作中蕴含着重要的思想资源，值得我们去挖掘、提炼。我在浙江大学跨学科社会科学研究中心的主要工作也就是从18世纪英国经验主义、情感主义理论出发，重新梳理社会科学思想和学科演进的理路，分析现代社会科学赖以产生的土壤，寻求丢失掉的传统，将其引入当下的时代，为社会科学的革新增添力量。

丁丁试图在学术传统资源和现代科学试验的基础上，把理性和情感统一起来。他提出了“情境理性”的范畴，我觉得这是挺有意思的。“情境理性”，有助于我们寻找一个打通中西学术基本范式的可能进路。在某种意义上，基于“情境理性”的秩序，或者是基于“情境理性”的法治社会如何演生，或许会成为中国社会科学当代最重大的本土化问题。下面，我把话筒交给叶老师。

三

叶　航：为了给大家作这个报告，汪老师准备了100多篇最新、最

前沿的文献。昨天，我们与研究生有一个小范围的座谈，发现没有时间把这些文献全部介绍给大家。好在，我们浙江大学跨学科社会科学研究中心跟踪这一方向国际前沿的研究已经很多年了，可以筛选出其中最重要的介绍给大家。这么多文献中，最有代表性的，实际上有两篇。

第一篇是《强互惠的演化——异质人群中的合作》[3]，发表在2004年2月美国《理论生物学》杂志上，这是一本引领国际生物学发展趋势的杂志。这篇文章的作者萨缪·鲍尔斯和赫伯特·金迪斯是桑塔费学派著名的经济学家。经济学家为什么要到生物学杂志上去发文章？因为现在社会科学前沿研究的跨学科倾向非常明显，而跨学科研究往往涉及经济学和生物学的结合。因为这两个学科都研究人的行为，从某种角度也研究人性。

这篇文章事实上是一个计算机仿真实验报告。它要解决的，就是刚才汪丁丁老师、罗卫东老师都提到的问题，即最经典的博弈案例“囚徒困境”中的合作问题。传统的主流经济学是建立在理性假设基础上的。你必须假设人是理性的，才能进一步推演出消费行为最大化和生产行为最大化。理性使经济行为变得有效率，从而实现帕累托最优。但在“囚徒困境”中，理性和效率是矛盾的。正因为人是理性的，怕别人背叛，最后导致了纳什均衡的非合作解。这意味着什么？意味着主流经济学逻辑体系不合法。为什么？因为你的两个假设前提，理性与效率发生了冲突，无法自洽。

上世纪50年代以后，许多经济学家试图解决这一难题。最初的思路是把单次囚徒困境看成重复博弈的一个子博弈，这样也许可以推出合作解，但也不能适用于所有场合。不过，这里仍然有问题。因为对主流经济学提出挑战的是单次囚徒困境博弈，你用重复囚徒困境博弈解决这个问题，在逻辑上是不是具有合法性？这样做事实上已经把前提改变了。我介绍的第一篇文献，就是为了解决单次囚徒困境条件下合作产生的问题。

刚才提到，这篇文献是一个计算机仿真实验的报告。仿真的环境

是距今20万年以前更新世晚期的人类狩猎—采集社会，仿真条件是严格按照古人类学已经确认的事实设置的。比如那时人类社会没有权威、国家政权，连酋长、宗教和巫师都没有；又比如那时人们采取的生产和消费方式是共同劳动和食物分享，而且也不储存食物。这样的条件一共有八个，都是严格按照考古学、人类学、历史学已有的证据设定的。在这样的社会中，人类要共同劳动，就需要合作，而合作就会碰到囚徒困境问题。

比如原始人一起出去狩猎，面对一只猛犸象，其他人都冲锋陷阵，有一个人躲在后面，但分享食物的时候他却出现了，这就是搭便车。显然，这种行为的生存适应性比冲锋陷阵要高。于是，从进化论角度看，不管这种差别最初多么微小，经过几百万年的进化，适应性高的行为会在一个生物群体中扩散开来，成为主导的行为模式。因此，英国生物学家道金斯说，如果你对生物学有所了解的话，就会得出一个必然的结论，所有从进化而来的东西都是自私的，包括人在内。[4] 这个结论事实上和主流经济学对人性的假设非常一致，我把它叫做“道金斯迷信”。但是桑塔费学派的经济学家却对这个结论提出了挑战。

根据鲍尔斯和金迪斯的计算机仿真，一个完全自私的人类族群，由于无法建立稳定的合作秩序，最终会趋于灭亡。合作秩序是怎样建立起来的呢？必须依靠一种被桑塔费学派称为“强互惠”的行为，即“Strong Reciprocity”。所谓强互惠行为，就是我首先和别人合作，如果对方背叛合作，哪怕这种背叛不是针对我，我也要进行惩罚，甚至不惜花费巨大的个人成本。在桑塔费学派的术语里，这种行为也被称为“Altruistic Punishment”，即“利他惩罚”。按照丁丁的说法，这是“见义勇为”，是“路见不平，拔刀相助”。事实上，这就是我们人类所特有的“正义感”。根据计算机仿真，只有当一个人类族群演化出这种行为后，才能建立起稳定的合作秩序。

我要介绍的第二篇文献，是《利他惩罚的神经基础》。[5] 发表在2004年8月的《科学》杂志上，是这期杂志的封面文章。这篇文献是

一篇脑科学的实验报告，是接着金迪斯、鲍尔斯所作的进一步研究。它要解决的问题是，如果强互惠行为或利他惩罚在人类合作秩序的建立过程中具有这么重要的作用，那么驱动这种行为的机制是什么？因为这种行为不同于自私行为，它无法给你带来利益上的激励。如果一种行为没有激励，它的动机是什么？这篇文献的通信作者是瑞士苏黎世大学国家经济实验室主任恩斯特·费尔博士，也是一位非常著名的桑塔费学派的经济学家。文章一开始提出了一个假设：如果强互惠行为或利他惩罚无法从外界获得直接激励，那么只有一种可能，就是行为者能够通过这种行为本身获得满足。也就是说，这种行为是依靠自激励机制实现的。

事实上，人和动物的许多行为都是依靠自激励机制实现的。脑科学已经证实，对高等动物来说，启动这类行为的机制是由中脑系统的尾核和壳核来执行的。比如我们人类和许多动物的成瘾性行为，像烟瘾、毒瘾和酒瘾等等，都涉及这一脑区。因此，这一脑区在医学上也称为“鸦片报偿区”。费尔博士猜测，如果强互惠行为依赖这种自激励机制，那么做出这种行为时，人脑的这个部位就会被激活，而且行为的强弱应该与其活跃程度正相关。于是，他们设计了一系列实验场景来激发人们的利他惩罚行为，并通过 PET 即正电子发射 X 射线断层扫描技术（Positron Emission Tomography）对脑神经系统进行观察。实验结果证实了这个大胆的推断。

实验结果显示，在预期的五个场合，与激励相关的脑区均被激活。尾核和壳核的血流峰值显示，其活跃程度远远超过平均水平，这时受试者表现出强烈的惩罚愿望并通过惩罚行为获得较高的满足。实验报告认为，社会偏好模型所定义的效用函数应该包含对违反公正和合作规范的惩罚愿望，这些模型可以比经济学传统的自利模型更好地解释人类的实际行为。强互惠或利他惩罚既不是一种像消化食物那样的自动机能，也不是一种基于深思熟虑、有明确目标导向的理性行为。这种依靠愿望诱导的激励机制说明，人们可以从这种行为本身获得满足。大

多数人在发现那些违反社会规范的行为未得到惩罚时会感到不舒服，而一旦公正得以建立，他们就会感到轻松和满意。

对人和其他高等动物来说，中脑系统是主管情感的脑区。在解剖学上，中脑也叫哺乳动物脑。它意味着，从古人类学和进化论的角度看，这一脑区在哺乳动物的时候就已经形成。人类的大脑皮层是在后来的长期进化中逐步形成的，覆盖在中脑系统上面。中脑所激发出来的行为主要是情感型的行为。因此，中脑比大脑更远古，情感比理智更远古。为什么会这样？因为早期的动物没有大脑，很多具有重大生存价值的行为，无法通过理性思维来实现。于是，就像今天的计算机芯片，内置了一个已经设计好的程序，一旦碰到相应的命令，这个程序就会自动执行。情感对我们来说无非就是这样一种内置的程序，这个程序在一定条件下，会让你自动执行某些动作，而无需理性的推断。

上面两篇文献代表着今天经济学与社会科学研究的前沿方向。国内的主流经济学是改革开放以后才引进的，我们曾经跟着西方人，鹦鹉学舌地告诉大家“人都是自私的”。但现在西方人已经走到前面去了，国内许多经济学家还在那里讲经济学不要讲道德、经济学讲道德是“狗拿耗子”。西方经济学家已经意识到这个问题，道德对市场经济来说不是可有可无的。缺失了道德维度，效率问题就没有办法真正解决。这些前沿研究对经济学乃至整个哲学社会科学都具有非常深刻、非常深远的意义，大致可以体现在以下六个方面。

第一，对“道金斯迷信”提出了挑战。道金斯认为，凡是进化而来的东西，其天性就是自私的。生物学家的这种看法，不是没有一点道理。随着现代基因技术和遗传科学的发展，所有实证研究似乎都证明了，生物进化必须通过个体的基因介质才能实现。两种不同的生物性状，比如A与B，假如A的遗传频率比B高，哪怕这种遗传优势微乎其微，也可能对生物进化产生重大影响。根据生物学家计算，某种生物性状只要有0.001的遗传优势，即使1年繁殖1次，经过23 400年也足以改变这个物种。[6] 从这点出发，当代生物学家事实上否定了生

物的利他主义行为。因为无论亲缘利他还是互惠利他，从基因层面都体现了一种自私性或利己性。由于利己行为的生存适应性大于利他行为，不管这种差别在初始状态多么微小，经过千百万年的自然选择，后者也会被无情淘汰。

但现在我们能够自信地指出，生物学家的上述看法是错误的，"道金斯迷信"可能来源于一个长期的误导和偏见。为什么？因为这一结论成立的前提，是孤立地考察利他行为与利己行为对适应性的贡献。我们认为，这个前提是不正确的。生物适应性是一个全面、综合的评价体系，它不可能被某个单一的事件决定。具体地说，一个利他者的生存适应性不仅取决于他与自私者的个别交往，还取决于他与其他利他者的交往；由于这些交往更容易达成合作从而使双方享受到合作剩余，因此只要这个剩余足够大，就能弥补利他者损失的进化优势。同样道理，一个自私者的生存适应性不仅取决于他与利他者的个别交往，而且还取决于他与其他自私者的交往；由于这些交往很难达成合作从而使双方无法享受合作剩余，如果这种损失足够大，就会使自私者攫取的进化优势损失殆尽。

如果不是孤立地考察利他行为与利己行为，而是在合作及合作剩余的框架中对生物个体的适应性进行全面评价，即使自然选择的基本单位是生物个体或个体的基因，利他行为也能够通过整体间的补偿机制获得相对的进化优势。因此，我们的结论与道金斯等主流生物学家的结论大相径庭——自私并不是人类惟一的天性！经过自然选择和进化产生的人类心智与人类行为，不仅与自利心相容，而且也与利他心相容。

第二，为解决单次囚徒困境博弈中的合作问题提供了新的思路。这个思路很简洁：突变产生出有利于合作的行为，合作导致了合作剩余；合作剩余增加了这类行为的生存适应性，从而有利于这种行为被自然和环境所选择，使行为人在进化过程中形成一种稳定的道德偏好；而道德偏好一经形成，就突破了博弈者原有的策略集合，打破了完全按自利原则推演出来的"纳什均衡"。就这么简单！丁丁前面提到的博弈

实验，在具有同情心的条件下，囚徒困境完全可以有合作解。其实，我猜测可能不仅仅是同情心，还有诸如正义感、愧疚感、宗教信仰等社会性情感，都可以在囚徒困境条件下诱导出合作行为。

第三，有助于我们重新认识道德的起源、道德的本质，甚至整个道德哲学。我们知道，传统的道德哲学中，有两大流派。一个是所谓的“义务论”，代表人物是康德；另一个是所谓的“后果论”，即功利主义的道德哲学，代表人物主要有边沁、休谟，也包括斯密。按照义务论的看法，道德行为不允许有任何功利的考虑。比如我做好事是为了大家的认同，或者不希望被人指责。康德说，这已经不是道德行为了。因为你在绝对的“善”之外，还有其他非善的目的。而后果论的看法刚好相反，它认为一种行为只有带来好的结果，才可能是道德的。道德哲学的这两大流派已经争论了二三百年，至今也没有结果。

但在我们现在这个理论框架下，义务论与后果论是可以统一的。从生物进化的角度看，道德的产生肯定是具有效率的事件，因为它是维护合作秩序不可缺少的要素。从这点看，道德具有后果论的功利性。社会生物学创始人威尔逊早就说过：“人在过去、现在和将来正是用它来保持人类遗传物质的完整无损，除此之外，道德并没有其他可以证明的最终功能。”[7] 但对每一个人来说，道德偏好一旦产生以后，你作出的道德选择，就没有任何功利目的了。因为，如果还能找到其他目的，我们终究可以把它归为某一类行为。就像费尔博士的脑科学实验告诉我们的，这种行为无需从外界获得激励，你必须对这种行为本身感到满意。我以为，这个实验正是用现代科学手段，揭示了康德“道德律令”的真谛。

第四，有助于我们重新认识某些重大的哲学范畴和哲学争论。比如，“实然”和“应然”。休谟最早区分了这两个不同的哲学范畴。所谓“实然”，指事物本身是什么。所谓“应然”，指我们应该怎么做。休谟认为，这是两个完全不同的研究范式。但在我们现在这个理论框架中，“实然”和“应然”之间没有不可逾越的鸿沟。我们看到，

像正义感、同情心、道德这些原本认为是“应然”的东西，事实上都有它进化的依据，可以被科学地分析，因此也是一个“实然”的过程。当然，并不是说“应然”就此消解、消失了。“应然”还是“应然”，仍然有“我们应该怎么做”的范式。只不过，在我们思考“应该怎么做”的同时，我们还要追问“为什么我们应该这么做”。也就是说，我们必须对“应然”本身作出“实然”的解释。另外，还有对“正义”的重新认识。由于时间关系，我就不再展开了。

第五，对情感因素在认知过程和决策过程中的地位与作用的重新认识。刚才，丁丁老师和罗老师都提到，经济学自认为是一门关于理性的学问，把情感因素完全拒绝在决策过程之外。而且我们还知道，从古希腊开始，哲学家就喜欢把情感和理性对立起来。好像情感只是兽性的延续，只有理性才是人性的昭显。因此，理性的一个重要表现就是对情感的纠正，或者是对情感的克制。当然，也有不同的意见。比如，休谟就认为“理性是激情的奴隶”。但不管怎么说，情感和理性，在传统范式中总是处于对立状态。但在我们现在这个理论框架中，情感与理性事实上是人类认知过程不可缺少的两个方面。正如丁丁指出的，在中国的文化传统中，“情”与“理”从来就是一个不可分割的整体。现在，西方人自己也开始反思这个问题。脑科学和认知科学的前沿研究已经揭示出，任何思维过程和决策过程事实上都不可能是单一维度的。回溯到上世纪 80 年代，心理学家也早就揭示过，一个人仅凭“智商”是不行的，一个健全的人还需要有“情商”。“情商”就是由情感激发的人类认知功能，像我们刚才讲过的那些在人类合作秩序建立过程中非常重要的品质，比如同情心、正义感、道德感等等，事实上都可以看作“情商”一类的东西。这些曾经使我们这个物种成功演化的优秀品质和善良情感，无论过去、现在和将来都是人类最可宝贵的财富。

最后，第六，有助于我们更深刻地认识制度演化的内在逻辑。我们认为，超越囚徒困境中个体理性的局限，谋求合作和合作剩余，可能是我们人类行为、人类心智与人类社会包括人类文化与人类制度共生演

化的最终原因。建立一个更完善、更有效率的合作秩序，也许是我们这个物种在生存竞争中的最大优势。在人类漫长的演化历史中，最初的合作秩序是通过自然选择建立的，即自然选择的压力迫使人类进化出有利于合作的偏好，我们把这一阶段称作“自然为人类立法”。随着生产能力的提高，自然施加于人类的选择压力开始减轻，合作秩序不得不通过其他手段来维护，强互惠者个人实施的利他惩罚就是其中之一，我们把这一阶段称作“个人为社会立法”。最后，在近现代社会，工业革命带来的分工使人类合作的规模达到前所未有的程度，合作秩序的维护必须依赖一个建立在民主基础上的现代司法制度，于是我们把这一阶段称作“社会为个人立法”。

汪丁丁:我为叶航老师补充一句。关于第六条，哈耶克很早就已经开始论证，现代社会是人类合作秩序不断扩展的结果。这就是所谓哈耶克的扩展秩序理念。我们今天可以从新的高度重新认识这个问题，尤其是哈耶克的扩展秩序与当代中国社会转型期的联系。回到休谟和斯密时代，他们对合作秩序也有类似看法。比如罗老师前面介绍的斯密的同情心理论，以及从“同情共感”出发的元心理学理论，还有休谟提出的“元美德”理论，其目的都是为了论证它们在市场制度演化过程中的重要作用。从同情心到对快乐的同情共感，就可以有我们人类仁慈感的出现，产生好施乐善的行为；从同情心到对痛苦与不幸的同情共感，就有了我们人类正义感的出现，产生嫉恶如仇的行为。这些都是休谟在《道德原则研究》里面说过的。

从那个时候开始到现在，中国人终于碰到了休谟问题。什么问题？我们没有上帝面前人人平等的法制，因此中国的市场经济就会遇到巨大的困难。所以，我们十几年来一直呼吁市场经济要有道德的基础。如果你既没有敬畏上帝的神学传统，又丢失了自己的道德传统，你所看到的市场经济就是肆无忌惮的贪污腐败、肆无忌惮的草菅人命、肆无忌惮的掠夺，以及目无法制。在这种境况下，高效率的合作秩序何以可能？

由于分工限制，我们浙江大学跨学科社会科学研究中心只能集中注意力进行纯粹的理论研究，但这并不是说我们不关注现实。事实上，我们对理论研究的偏好，正是源于我们对现实的感悟和忧虑。我希望在座的同学们——年轻的学子们，更多关注我们报告中提到的问题，因为它既代表着主流经济学和整个社会科学未来发展的方向，也是解决我们中国当下问题的良药。

注释：

[1] Marshall, *Principles of Economics* (London: The Macmillan Company, 1890 (1938)).

[2] David Sally, On Sympathy and Games, *Journal of Economic Behavior & Organization*, Vol. 44(2001), 1—30.

[3] Gintis & Bowles, The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations, *Theor. Popul. Biol*. 65, 1, 2004. 中译本见：金迪斯，鲍尔斯等著，2005，《走向统一的社会科学》，上海人民出版社，第72—99页。

[4] Dawkins, The Selfish Gene (Oxford Press, 1976).

[5] Fehr et al., The Neural Basis of Altruistic Punishment, *Science*, Vol. 305, 2004.

[6] 陈阅增，1997，《普通生物学》，高等教育出版社。

[7] Wilson, *On Human Nature* (Harvard, University Press, 1978).

导读二：

作为内生偏好的利他行为及其经济学意义*

叶　航　汪丁丁　罗卫东

1　传统：一个经济学与生物学的回顾

1.1　经济人假设与理性人假设

现代主流经济学把自利作为人类行为的基本前提，从而本质上排斥了利他行为[1]对经济研究的意义（张五常，2001；田国强，2005）。这一传统可以追溯到所谓的经济人假设。亚当·斯密在《国富论》中把追求利润最大化的个人确立为经济分析的出发点，为

* 本文发表于《经济研究》2005年第8期。本文得益于浙江大学跨学科社会科学研究中心（ICSS）与美国桑塔费研究院（SFI）的学术交流，特别是鲍尔斯教授、金迪斯教授与我们的深入讨论。中国社科院经济研究所杨春学研究员对经济人和利他主义经济学的系统研究（1998，2001），复旦大学经济学院韦森教授对经济学与伦理学关系的深刻思考（2002），也使我们获益匪浅。在此，谨向以上同行和朋友表示由衷感谢。此外，作者还要感谢教育部"语言与认知研究"国家哲学社会科学创新基地和浙江大学"强所计划"对本项研究的支持。

新古典经济学和现代主流经济学奠定了分析生产者行为的基本范式（Smith，1776）。19 世纪 50—70 年代的边际革命把追求效用最大化的个人确立为经济分析的另一个出发点，为新古典经济学和现代主流经济学奠定了分析消费者行为的基本范式（Gossen，1854；Jevons，1871；Menger，1871；Walras，1874）。这两个范式内在地统一于追求自身利益最大化，因此帕累托把具有这种行为倾向的人概括为“经济人”，并认为它是全部经济分析的前提假设（Pareto，1896）。由于这个假设隐含着一种对人性自私的肯定，一经面世就引发了众多批评，其中也包括来自经济学内部的批评。

20 世纪 20 年代以后，经济人假设逐步被理性人假设取代，这种取代主要基于两个原因：第一，来自外部和内部的持续批评，使许多经济学家——也许他们并不赞同这些批评，但为了避免怀疑和争论——不得不在表述时使用一些更抽象的术语，比如最大化行为、最优决策、理性选择等，从而导致了“理性人”这一概念的流行（杨春学，1998）；第二，20 世纪 30—50 年代，萨缪尔森出于经济学数理化的需要，对许多传统经济学概念进行了重新表述，而效用的重新表述导致对理性和理性人的重新定义，并最终确立了它在现代经济学中的地位。根据现代经济学的解释，效用是偏好的函数，用偏好定义理性，只需满足完备性和传递性两条假定（Mas-Colell， Whinston & Green，1995）。而所谓理性人，简而言之就是约束条件下最大化自身偏好的人。

新的定义为经济学提供了一种“去伦理化”的可能。虽然大部分经济学家仍然在自利范围内使用这个术语，但经济学对偏好的定义事实上不依赖偏好的伦理取向。换言之，经济学所谓的偏好，既可以包括利己偏好也可以包括利他偏好。正是在这种理论背景下，贝克尔开创性地用理性选择模型对利他偏好作出了解释（Becker，1976），从而使利他行为逐步进入主流经济学的研究视野。但早期的研究主要是在外生给定利他偏好的前提下进行的，即“只需假设利他主义者所要最大化的不仅仅是他们自己的个人福利，还有他们所关心的某些其他人的福利”（杨春

学，2001），就可以对诸如自愿献血、慈善捐款和非营利组织等利他行为作出标准的经济学分析（Sugden，1982；Collard，1983）。

可是，一旦经济学试图将利他偏好内生化，需要解释利他偏好的形成原因时，经济学家就发现他们将面临和生物学家同样的问题："减少个人适应性的利他行为如何能够通过自然选择而得以进化？"（Wilson，1975）于是，很多经济学家试图从伦理学、社会学、心理学或者人类学中寻求答案，把利他偏好的原因归结为道德规范、文化教育，甚至宗教信仰（Cavalli-Sforza et al.，1981；Lumsden et al.，1981；Boyd et al.，1985），"然而，不幸的是，这些学科尚未形成更为系统或可资利用的偏好知识。"（Becker，1976）事实上，即便经济学家可以从这些途径中找到答案，对经济学来说，可能仍然于事无补。利他偏好的内生化，需要经济学家在经济学的逻辑体系和框架中寻求解释。

贝克尔是最早尝试这一努力的经济学家。他认为，如果把利他行为看成适应性的生产过程，利他主义者最大化自己和受惠者适应性的总和，利他行为的均衡是施予者的边际适应性等于受惠者的边际适应性，因此，"利他主义并非像以往定义的那样必然会减少个人适应性"（Becker，1976）。这个解释虽然勉强，但它的结论并非没有道理（本文第三部分将证明这点）。西蒙则把社会奖赏作为一种激励机制引入经济学对利他主义的分析，他认为，如果这种奖赏大于利他者相应减少的生存适应性，"利他主义就会逐步在人口中占据支配地位。"（Simon，1982）但西蒙没有说明这种激励机制产生的原因，因此这个解释不彻底，只是用一个外生变量替代了另一个外生变量。伯格斯特朗和斯塔克证明了亲属或邻居之间在单次囚徒困境博弈中可以产生合作，并推论合作剩余有利于利他主义的进化，因为"基因遗传是一个迟钝的过程，一般不会孤立地对个人发挥作用；那些具有合作倾向或是继承了有利于这一倾向基因的人，更可能比其他人享受到合作带来的利益。"（Bergstrom & Stark，1992）这个观点已经接近（本文第二部分将要介绍的）桑塔费学派的最新认识。

另一方面，许多经济学家，其中包括 Hirshleifer(1977)、Lindbeck 和 Weibull （1977）、Collard（1978）、Nakayama（1981）、Arrow (1982)、Hammond(1987)等，坚持认为利他主义对经济学是一个多余的假设，它的存在可能导致经济活动的帕累托无效率。比如，有名的“先走悖论”，即如果每人都坚持对方先走，结果无人能通过一道大门；还有所谓的“萨玛利亚人困境”（Samaritan's dilemma），即对未来援助的预期可能诱使人们过度消费，通过故意恶化自己处境的方法来获得更多资源；等等（更详尽的介绍参见杨春学，2001）。虽然利他主义行为在慈善事业、非盈利组织、公共物品领域的作用是有目共睹和显而易见的，但大部分经济学家仍然用沉默表明了他们在这个问题上的主流立场。

1.2 群体选择理论与个体选择理论

生物学与经济学内在的逻辑相当接近，达尔文和华莱士都是受经济学家马尔萨斯的启发，才萌发了“物竞天择，适者生存”这一进化论的基本思想（Bowler，1984）。达尔文在自传里写道：“1838 年 10 月，……我偶尔翻阅了马尔萨斯的《人口论》。当时，我脑海里已经孕育了生存斗争的思想。通过对动植物生活习性的观察，我发现这种斗争无处不在。马尔萨斯的著作立刻吸引了我，在有限的空间里，只有适者才能够继续存在，而不适者势必遭到淘汰，结果形成新的物种。于是，我终于找到了一种继续工作的理论基础。”（Darwin，1887）有人对《美国经济评论》和《美国博物学家》刊载的文章进行过比较，结果发现这两门学科包含的内在逻辑惊人地相似，所有生命体的行为看上去总好像设法使某一目标函数最大化，而典型的论文都是运用优化方法预测某种现象，然后再作出统计校验（Tullock，1983）。

达尔文的进化论曾经遭到许多曲解，以至于“达尔文主义”一度成了冷酷无情的代名词（Wright，1994）。但客观地说，达尔文并没有像以后的生物学家那样，把《物种起源》揭示的逻辑始终如一地贯彻到人

类和人类天性上去。1871年，《物种起源》出版12年后，达尔文出版了《人类的由来》，在解释人类道德感时，他说“道德水准较高，多数人奉行道德规范的部落，绝对比其他部落更为有利。无疑，一个部落若有许多热爱群体、忠于群体、服从群体，既勇敢又体恤他人，随时准备互相支援并为共同利益自我牺牲的人，必能战胜其他大多数部落；这便是天择。”（Darwin，1871）虽然有人怀疑这是达尔文屈服于维多利亚时代虚伪道德传统的违心之言，但这个思想毕竟为群体选择理论提供了依据（Wright，1994）。

直至20世纪60年代以前，也就是《物种起源》出版后的100年间，群体选择理论事实上是大多数生物学家关于进化的主流范式（Martin，2001），其追随者包括1973年诺贝尔生物医学奖获得者劳伦兹和丁伯根、美国著名生态学家埃默森、英国著名生物学家爱德华兹以及社会生物学的创始人威尔逊等。该理论的拥戴者认为，自然选择是在生物种群层次上实现的，当生物个体的利他行为有利于种群利益时，这种行为就可能随种群利益的最大化而得以保存。当面临巨大灾变或是种群之间的生存竞争时，一个存在着利他主义的生物种群与一个完全缺乏这种献身精神的生物种群相比，具有更大的生存适应性。因此，利他行为可以伴随着种群的胜利而成功演化（Edwards，1962；Wilson，1975）。这个思想和100年前达尔文在《人类的由来》中所表达的思想如出一辙。

上述传统在20世纪60年代以后发生了重大转向，肇始者是美国纽约州立大学生态学家威廉斯。他在《适应与自然选择》一书中声称，自然选择只能作用于生物个体，这是对达尔文进化思想的捍卫（Wiliiams，1962）。在威廉斯带领下，生物学内部展开了一场对群体选择理论的清算，并逐步使个体选择理论占据了主流地位。1964年，英国生物学家汉密尔顿首创了进化论史上的一个重要概念“亲缘选择”，成功地从个体角度解释了生物世界普遍存在的亲缘利他行为（Hamilton，1964）。1971年，哈佛大学进化论教授特里弗斯借助博

弈论解释了生物个体之间的互惠利他行为（Trivers，1971）。个体选择理论真正大行其道，也许得归功于牛津大学著名生物学家道金斯，他1976年出版《自私的基因》一书，使这个理论走出生物学家的书斋，成为一般公众的常识。

道金斯认为，“自然选择的基本单位，也就是自我利益的基本单位，既不是物种，也不是群体。从严格意义来说，甚至也不是个体，而是基因这一基本的遗传单位。”（Dawkins，1976）两种不同的生物性状，比如A与B，假如A的遗传频率比B高，哪怕这种遗传优势微乎其微，也可能对生物进化产生重大影响。根据生物学家计算，某种生物性状只要有0.001的遗传优势，即使1年繁殖1次，经过23 400年就足以改变这个物种（陈阅增，1997）。从这点出发，当代生物学家事实上否定了生物的利他主义行为。因为无论亲缘利他还是互惠利他，从基因层面都体现了一种自私性或利己性。由于利己行为的生存适应性大于利他行为，如上所述，不管这种差别在初始状态多么微小，经过千百万年的自然选择，后者也会被无情淘汰。以至于道金斯斩钉截铁地说，“如果你认真研究了自然选择的方式，你就会得出结论，凡是经过进化而产生的任何东西，都应该是自私的，”“对整个物种来说，‘普遍的爱’和‘共同的利益’在进化论上简直是毫无意义的概念。”（Dawkins，1976）

从演化均衡的角度看，道金斯说，即便一开始存在一个没有叛逆者的利他主义群体，我们也很难阻止自私个体的侵入，因为不能保证不会由突变而产生一个自私的个体；只要产生了一个叛逆者，它不但拒绝作出任何牺牲，而且还会利用别人的牺牲为自己牟利；按照定义，它就会比其他成员有更大的机会生存下来并繁殖自己的后代，而这些后代都会继承其自私的特征；这样的自然选择经过几十或几百代以后，利他的个体就将被自私的个体湮没，利他的群体与自私的群体就没有办法分辨了（Dawkins，1976）。因此，道金斯认为利他行为不是一个“演化稳定策略”（evolutionarily stable strategy，ESS），因为它无法抵御自私行

为的侵入。反之，包括亲缘利他和互惠利他在内的利己行为却具有很强的稳定性。

道金斯强调基因自私性时虽然也考虑到了人类的道德问题，但他认为，道德必须从外部强加在一个本质自私的人身上，“我们能做的只是尽最大可能来宣扬慷慨大度和克己利人的精神”，因此“你不要指望从人的天性中得到任何帮助，因为我们天生是自私的”。(Dawkins，1976) 根据阿莱克什达的理解——他是上述生物学传统中最有影响力的伦理学家——甚至社会道德也只能从表面上超越自私，他断言“只有把社会看作一个追求各自利益的个人集合，我们才能理解伦理、道德、人类行为和人类心理”。(Alexander，1987) 美国生物学家杰塞林甚至宣称，“如果不是感情用事，我们会发现没有任何迹象表明纯粹的慈善行为会改善我们对社会的看法，所谓的合作事实上只是机会主义和利用他人的结合体。”(Ghiselin，1974)

作为一种进化方式，群体选择面临的最大困难在于，随着现代基因技术和遗传科学的发展，所有实证研究似乎都证明了：生物进化必须通过个体的基因介质才能实现——有利于个体适应性的生物性状才会在遗传中得以保存和进化，与个体适应性无益或有害的生物性状最终都会在遗传中丢失和湮没。群体选择理论正是在这个关键问题上存在着一个致命的弱点：它无法解释能够给群体带来利益但却导致个体适应性降低的利他行为怎样才能在严酷的生存竞争中对利己行为保持相对的遗传优势，从而使自己得到进化。

2　前沿：来自桑塔费学派的最新看法

2.1　强互惠、利他惩罚及其演化均衡

与生物学家对人性的悲观判断相反，1990 年代以后，随着实验经

济学与行为博弈理论的发展，经济学家发现，人类相当一部分带有利他倾向的行为，无法用亲缘理论和互惠理论解释。其中最重要的是一种被桑塔费学派经济学家称为“强互惠”（Strong Reciprocity）的行为，这种行为的特征是：在团体中与别人合作，并不惜花费个人成本去惩罚那些破坏合作规范的人（哪怕这些破坏不是针对自己），甚至在预期这些成本得不到补偿的情况下也会这样做（Gintis，Bowles，Boyd，Fehr，2003）。强互惠能抑制团体中的背叛、逃避责任和搭便车行为，从而有效提高团体成员的福利水平。但实施这种行为却需要个人承担成本，并且不能从团体收益中得到额外补偿。从这点看，强互惠是一种明显具有正外部性的利他行为。因此，桑塔费学派也把这种行为称作“Altruistic Punishment”，即“利他惩罚”（Fehr et al.，2004）。

在经典的公共品博弈中，受试者得到允诺，只要向公共账户投入自己的钱币，每个人都将获得奖励。与任何一个公共品的生产一样，这个博弈的关键在于，即便你没有投入一分钱，也可以通过搭便车提高自己的福利。根据理性假设，该博弈的纳什均衡是所有博弈者都不向公共账户捐赠。但实际上，只有少数受试者符合这一推断。相关实验显示，最初几轮博弈中，捐赠的平均水平在40%到60%（每人持有的初始货币为20元）之间。随着博弈的进行，捐赠有所降低，最后一轮有73%（总数是1 042）的个体拒绝捐赠。这个结果与理性人假设相符，即博弈者在重复博弈的最后一轮倾向于背叛。但实验后的调查却出乎人们预料，当问及为什么减少捐赠或拒绝捐赠时，大部分人声称这样做是出自愤怒，是想通过这个自己拥有的惟一手段来惩罚那些搭便车者（Fehr & Schmidt，1999）。

受试者的解释是可信的吗？在一个新设计的公共品博弈中，受试者被允许对搭便车行为进行惩罚，即他可以要求罚没某个人的钱币，但行使这个权力必须支付一定的费用。这个实验的关键是，惩罚可以减少搭便车行为从而增加公共福利，但却需要个人支付成本。根据理性假设推断，又会产生第二种意义上的搭便车，即大家都希望别人来实施

惩罚，而自己坐享其成。但实验结果却显示，带有利他性的惩罚相当普遍，而且整体捐赠水平也因此明显提高。（Fehr & Gächter，2000）事实上，许多行为博弈实验，包括最后通牒博弈（Güth，1982；Blount，1995；Gintis，2003）、劳动市场博弈（Fehr，Gächter & Kirchsteiger，1997）、偷袭者博弈（Falk，2002）等都证实了这点。由于利他惩罚的存在，不但博弈者的策略集改变了，而且根据理性假设预测的博弈均衡也改变了。

桑塔费研究院经济研究所所长萨缪·鲍尔斯和赫伯特·金迪斯教授认为，人类行为具有的这种特征，可能是我们这个物种在漫长进化过程中形成的一种特定的行为模式。当严酷的生存竞争迫使人类把合作规模扩展到血亲关系以外，而普遍存在的单次囚徒困境又无法为互惠行为提供条件时，由基因突变产生的强互惠或利他惩罚，可以侵入完全自私的人类群体，从而有效维护族群内部的合作规范，显著提高族群的生存竞争能力。为了证实这个猜想，桑塔费研究院通过计算机仿真技术，模拟了距今10万—20万年以前狩猎—采集社会的人类生活，实验结果支持了这个假设。2004年2月，美国《理论生物学杂志》发表了鲍尔斯和金迪斯撰写的一篇重要论文《强互惠的演化：异质人群中的合作》（Bowles & Gintis，2004），详细介绍了这次实验的过程和结论。

计算机仿真的背景是更新世（Pleistocene）晚期人类狩猎采—集族群的生态与生活，仿真条件严格按照考古学和古人类学对这个社会已有的知识设定，这些条件包括：（1）族群规模不大，族群成员之间可以相互观摩和交往；（2）不存在社会权威，比如酋长、宗族或宗教领袖，社会规范的维护依赖个体的参与；（3）族群不是建立在亲缘关系基础上，不能用亲缘理论解释可能出现的利他行为；（4）个体之间的地位差异非常有限，主要根据行为特征而不是身份对族群成员进行分类；（5）分享是主要的消费方式，除非藏匿，否则个体或群体劳动获得的食物都在族群成员间平均分配；（6）个体不储存食物或积累资源，采取所谓“即时回报”（immediate return）的生产系统；（7）驱逐是族

群内部进行惩罚的主要形式，个体可以用逃离族群的方法躲避更为严厉的惩罚措施；（8）族群中个体的行为存在小概率变异的可能，即由一个正的突变率引导出不同的行为类型。

根据族群成员对待合作劳动的态度，把他们的行为分成三种基本类型：（1）自私者（Selfish），他们总是企图分享合作成果，而竭力逃避合作责任；（2）合作者（Cooperator），他们无条件提供合作劳动，但不会惩罚背叛者；（3）强互惠者（Reciprocator），他们与别人合作，并不惜花费个人成本惩罚违反合作规范的人。仿真动力学模型由 7 个相关方程组成：它们分别决定了个体繁殖率，行为突变率，合作劳动的成本与收益（其净值体现为个体生存适应度），惩罚和被惩罚的成本等重要参数。根据生物学和人类学知识外生给定：（1）生存适应度小于零即视为个体死亡；（2）规模小于 7 人视为族群灭绝。仿真的初始状态为：（1）20 个相对独立的原始族群，每个族群的规模为 20 人；（2）族群成员 100%都是自私者，即把生物学个体选择理论预测的结果作为仿真起点，然后检验其稳定性。下图是计算机仿真行经 3 000 代演化均衡的动态过程：

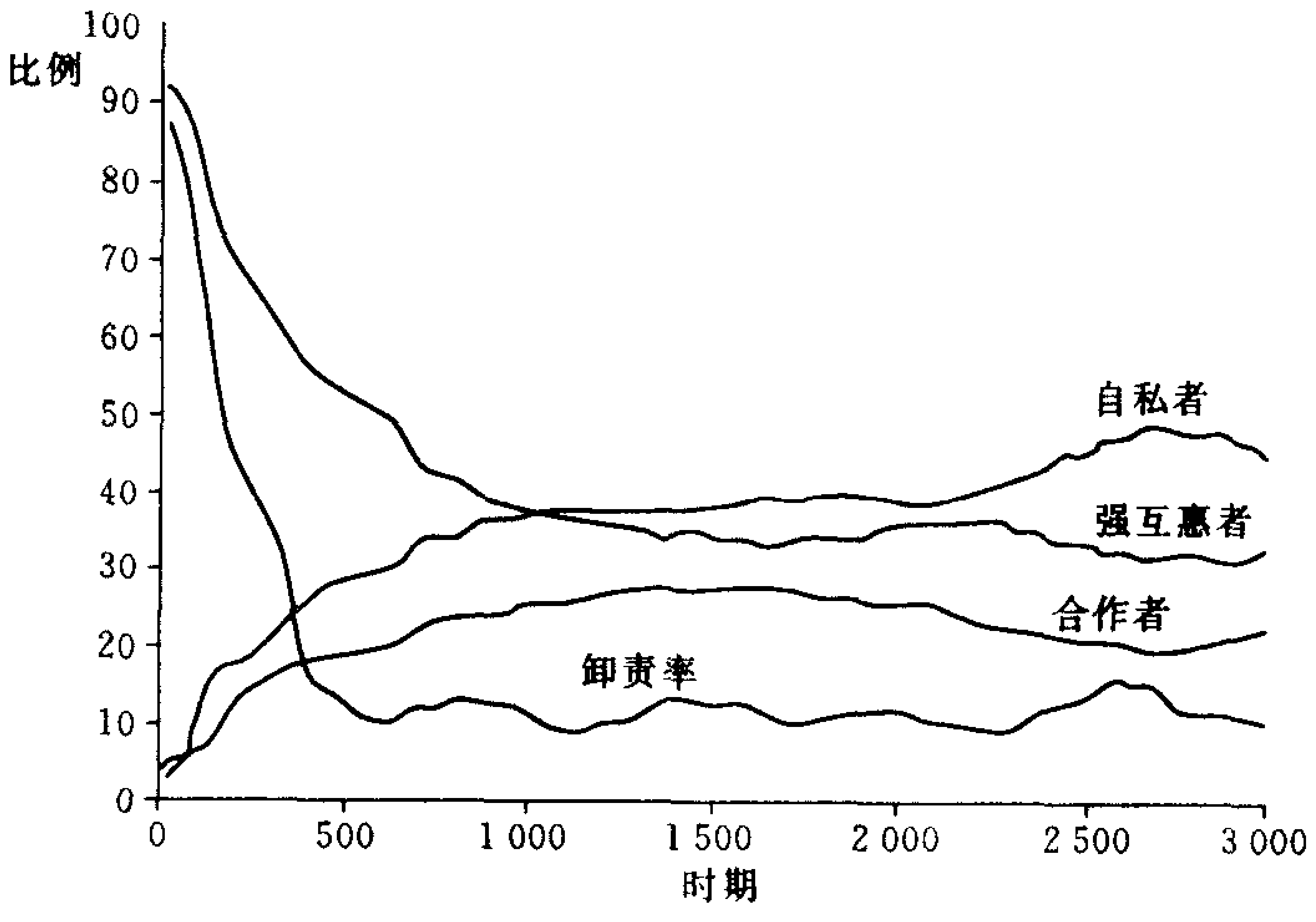

图 1

仿真结果显示：(1) 由突变产生较小数量的强互惠者可以侵入自私者人群，使群体内的合作行为与适应性维持在一个较高水平；(2) 只演化出合作者的群体是不稳定的，单纯的合作行为不具备生存优势，群体最终将回归初始状态；(3) 完全由自私者组成的群体，由于缺乏合作机制维持的适应性相对优势，最终将导致灭绝。演化均衡的动态过程显示：仿真的初始阶段，自私者占统治地位，逃避合作导致的卸责率 (Shirking Rate) 接近 100%；其后，随着强互惠者的出现，合作者的人数开始增加，平均卸责率则迅速降低；在大约 500 代左右，卸责率下降到 10%的水平，而强互惠者和合作者在群体中的比例继续上升；在其后大约 2 500 代内，群体中三种人群的比例及其平均卸责率基本维持在一个稳定水平，其均值为：自私者占 38.2%，合作者占 24.6%，强互惠者占 37.2%，平均卸责率为 11.1%。

如果把这一结果看作群体成员中三种行为发生的概率，则可以有一个更符合实际的结论：在上述条件下，通过演化而形成的人类行为大约有 38.2%的概率表现出自私倾向，24.6%的概率表现出单纯合作的倾向，37.2%概率表现出强互惠倾向；平均而言，每个人因机会主义充当搭便车者的可能性大约为 11.1%。

2.2 强互惠、利他惩罚及其激励机制

强互惠或利他惩罚是一种具有正外部性的利他行为，但这种行为的激励机制是什么？在得不到物质补偿的情况下，人们为什么不惜花费个人成本去惩罚那些违反合作规范的人？桑塔费学派的重要成员、苏黎世大学国家经济实验室主任恩斯特·费尔博士猜测，如果不能从外界得到必要的激励，强互惠者只能从利他惩罚行为本身获得预期的满足。为了证实这个假设，苏黎世大学国家经济实验室使用 PET 即正电子发射断层扫描技术 (Positron Emission Tomography) 对这一行为的脑神经系统进行了观察。相关研究表明，位于中脑系统的纹体 (striatum) 包括尾核与壳核的神经回路，是人类及灵长类动物整合激励信息与行为信

息的关键部位。如果利他惩罚的发生是惩罚者预期从惩罚行为本身得到满足，通过 PET 应该观察到这一脑区的激活，且惩罚行为的强弱与其活跃程度正相关。实验结果证实了这个大胆的推断。2004 年 8 月，《科学》杂志以封面文章的重要地位发表了有关这一实验的报告：《利他惩罚的神经基础》（Fehr et al.，2004）。

实验过程大致如下：两个受试者 A 和 B 为一组，每人都得到 10 单位初始货币；第一步，A 可以选择把自己的货币全部交给 B，如果 A 这样做，实验者就把 A 交给 B 的货币扩大 4 倍，即 B 可以得到 40 单位货币；第二步，B 决定是否把 50 单位货币（自己 10 单位加上被赠与的 40 单位）中的 50%回赠给 A，如果 B 不这样做，则信任他的 A 将分文不得；第三步，A 被赋予惩罚 B 的权利，即可以罚没 B 所拥有的货币。A 有一分钟时间思考是否实施惩罚，以及惩罚的数量。实验者在这一分钟内通过 PET 对 A 的大脑进行扫描，一共有 14 位经历了背叛的 A 被作为观察样本。

为了测试惩罚行为与大脑兴奋之间的关联，实验者设计了 4 个不同场景：（1）称为 IC（有意但有代价），即 B 有意滥用 A 的信任，但 A 惩罚 B 是有代价的；（2）称为 IF（有意但无代价），即 B 有意滥用 A 的信任，但 A 惩罚 B 是无代价的；（3）称为 IS（有意但象征性），即 B 有意滥用 A 的信任，但 A 对 B 的惩罚是象征性的，不能实质上减少 B 的货币；（4）NC（无意但有代价），即 B 的行为是随机的，比如通过骰子来决定（且 A 被事先告知），但 A 惩罚 B 仍然是有代价的。为了控制序列影响，四种情况出现的顺序是随机决定的。

根据验前的假设推断，（1）A 在 IF 和 IS 条件下都有惩罚 B 的愿望，因为 B 是故意的；但由于 IF 是实质性的惩罚，IS 是象征性的惩罚，因此后者的满意程度较小，激励相关脑区的活跃程度应该低于前者。（2）如果 IF 条件下的惩罚是令人满意的，受试者也会接受相应的惩罚成本，那么，在 IF 条件下激励相关脑区的高度活跃者，应该在 IC 条件下愿意为惩罚承担较高的成本。（3）由于 B 不需要为 NC 条件

下的行为负责，这样A就没有或者只有很微弱的惩罚愿望。因此，NC条件下激励相关脑区不会被激活或激活程度很低。

实验结果证明，A在IC、IF和IS三种情况下都显示出强烈的惩罚愿望，IF条件下全体受试者都对B实施了惩罚，IC条件下14个受试者有12个对B实施了惩罚，IS条件下14个受试者有6个对B实施了惩罚；与此形成鲜明对比的是，在NC条件下A几乎没有惩罚愿望，14个受试者只有3个惩罚了B，而且惩罚强度相当低。

实验结果显示，在预期的五个场合下，与激励相关的脑区均被激活。尾核血流峰值显示，在IC和IF条件下其活跃程度超过平均水平，这时受试者表现出强烈的惩罚愿望并通过惩罚行为获得较高的满足；在IS和NC条件下，其活跃水平低于平均水平，受试者要么不能满足惩罚愿望要么没有惩罚愿望。实验报告指出，尾核的兴奋值得高度关注，因为这一区域对行为激励具有显著作用。在大鼠的损伤性实验以及灵长类动物神经元实验的记录中，这一脑区与激励信息密切相关。人类尾核的神经成像研究也证明，该脑区的活跃与行为激励过程相关。此外，在诸如可卡因和尼古丁的强化刺激下，也发现了尾核的活跃。

实验报告认为，最新的社会偏好模型所定义的效用函数包含了对违反公正和合作规范的惩罚愿望，这些模型比经济学传统的自利模型更好地解释了人类的实际行为。利他惩罚行为既不是一种像消化食物那样的自动机能，也不是一种基于深思熟虑、有明确目标导向的行为。这种典型的依靠愿望诱导的激励机制说明，人们可以从这种行为本身获得满足。大多数人在发现那些违反社会规范的行为未得到惩罚时会感到不舒服，而一旦公正得以建立他们就会感到轻松和满意。

3　利他行为的演化均衡及其相关讨论

我们在本文第一部分曾经指出，从个体选择理论出发，当代生物学

家事实上否定了生物的利他主义行为。因为无论亲缘利他还是互惠利他，从基因层面都体现了一种自私性或利己性。由于利己行为的生存适应性大于利他行为，不管这种差别在初始状态多么微小，经过千百万年的自然选择，后者也会被无情淘汰。本文第二部分，我们介绍了桑塔费学派经济学家的最新研究成果。这些研究成果表明，人类相当一部分带有利他倾向的行为，无法被亲缘利他和互惠利他解释。当严酷的生存竞争迫使人类把合作规模扩展到血亲关系以外，而普遍存在的单次囚徒困境又无法为互惠行为提供条件时，以强互惠为特征的利他行为可以侵入完全自私的人类群体，使族群内部的合作劳动与相对适应性维持在一个较高水平。

经济学家的看法与生物学家的看法产生了严重分歧。对桑塔费学派的经济学家来说，他们今天仍然面临着100多年来群体选择理论面临的同样问题：利他行为演化均衡的微观基础是什么？导致自身适应性降低的利他行为怎样才能在严酷的生存竞争中对利己行为保持相对的遗传优势，从而使自己得到进化？我们以下的研究将证明，道金斯等主流生物学家的观点存在着明显的疏漏。个体选择理论断言利己行为比利他行为具有更大的适应性，也许产生于一个长期的偏见和误导。因为这一结论得以成立的前提，是孤立地考察利他行为与利己行为对适应性的贡献。我们认为，这个前提是不正确的。生物适应性是一个全面、综合的评价指标，它不应该而且也不可能被某个单一的事件或关系所决定。

具体地说，一个利他者的生存适应性不仅取决于他与自私者的个别交往，而且还取决于他与其他利他者的交往，由于这些交往更容易达成合作从而使双方享受到合作剩余，只要这个剩余足够大，就能弥补利他者损失的进化优势。同样道理，一个自私者的生存适应性不仅取决于他与利他者的个别交往，而且还取决于他与其他自私者的交往，由于这些交往很难达成合作从而使双方无法享受合作剩余，如果这种损失足够大，就会使自私者攫取的进化优势损失殆尽。我们可以通过以下例子证明这个观点：

表 1

	利他者	利己者
利他者	5，5	0，10
利己者	10，0	−2，−2

上表中，尽管利己者可以从利他者身上获取很大的利益（10，0 或 0，10），但如果合作行为为利他者带来的利益（5，5）远远超过利己者损失的合作剩余（−2，−2），利他者在进化过程中仍然具有相对的生存优势。

以 X 代表利己者的数量，以 Y 代表利他者的数量；把上表假定的损益当作生物个体不同境况下的生存适应性，则利己者的期望适应性 $EU_X=-2X+10Y$，利他者的期望适应性 $EU_Y=5Y$；当利他者与利己者之比 $Y/X=2/5$ 时，每个个体的生存适应性都是一样的（$-2X+10Y=5Y\rightarrow Y/X=2/5$）；如果利他者与利己者之比小于 2/5，利他者的适应性大于利己者，则利他者的数量将趋于增加；反之，如果利他者与利己者之比大于 2/5，利己者的适应性大于利他者，则利己者的数量将趋于增加。因此，演化均衡将使利他者和利己者的比例收敛于 2/5，即在上述条件下，该生物群体中的个体行为有 40%的概率表现出利他主义倾向，60%的概率表现出利己主义倾向。这两种行为都是"演化稳定策略"。在横轴为利他者与利己者的比例 Y/X，纵轴为期望生存适应性 EU 的坐标中，二者的演化均衡如下图所示：

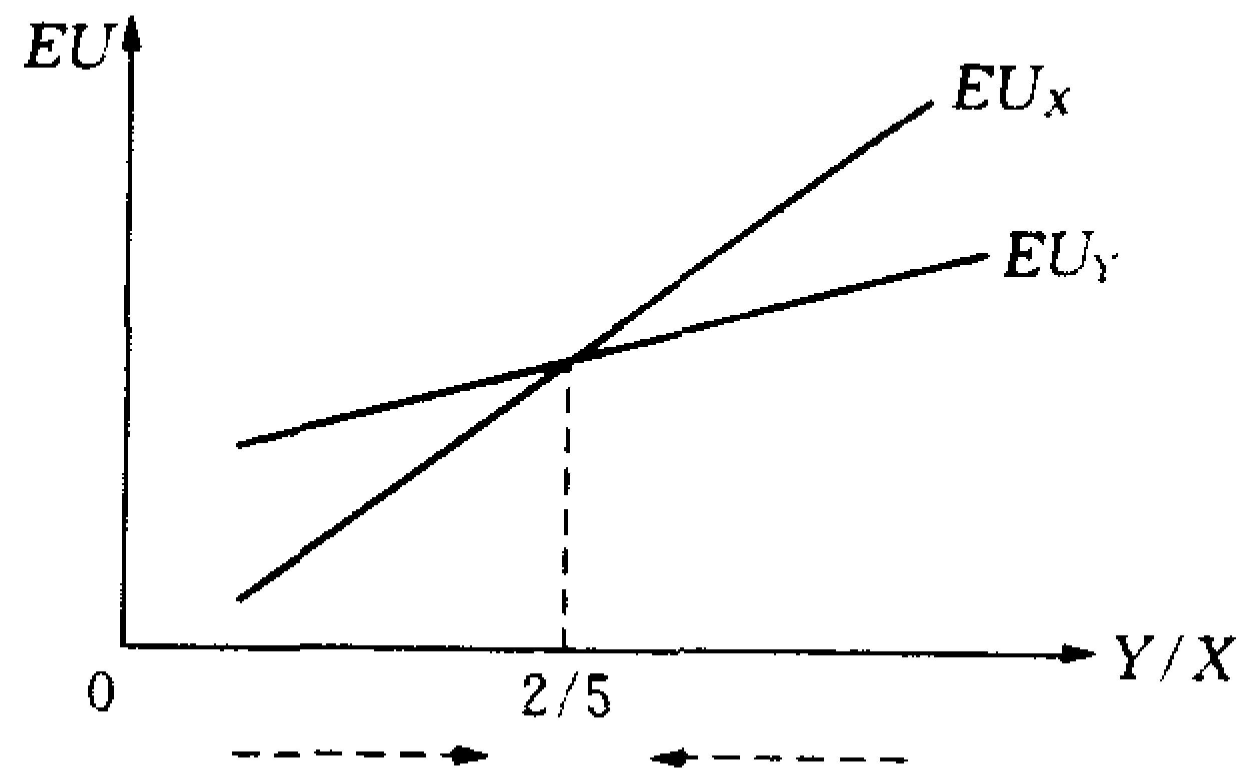

图 2

一般地，如果以 V 代表利他合作带来的全部收益，则 $V/2$ 表示两个利他者相遇时的平均收益，0 与 V 或 V 与 0 则表示利他者遭受背叛而利己者独占全部合作收益；如果以 C 代表合作剩余，则 $(V-C)/2$ 表示丧失合作剩余情况下每个利己者的平均收益；上述关系可以表示为：

表 2

	利他者	利己者
利他者	$V/2$，$V/2$	0，V
利己者	V，0	$(V-C)/2$，$(V-C)/2$

以 X 代表利己者的数量，以 Y 代表利他者的数量，则利己者和利他者的期望适应性分别为：

$$EU_X = VY + (V-C)X/2 \tag{1}$$

$$EU_Y = (V/2)Y \tag{2}$$

演化均衡 θ^* 为利他者与利己者的预期适应性相等，即：

$$\theta^* = EU_X = EU_Y \tag{3}$$

将（1）、（2）式代入有 $\theta^* = VY + (V-C)X/2 = (V/2)Y$，化简后有：

$$\theta^* = Y/X = (C-V)/V \tag{4}$$

如果考虑可能出现的非线性状态，在横轴为利他者与利己者的比例 Y/X，纵轴为期望生存适应性 EU 的坐标中，我们可以把上述演化均衡表示为：

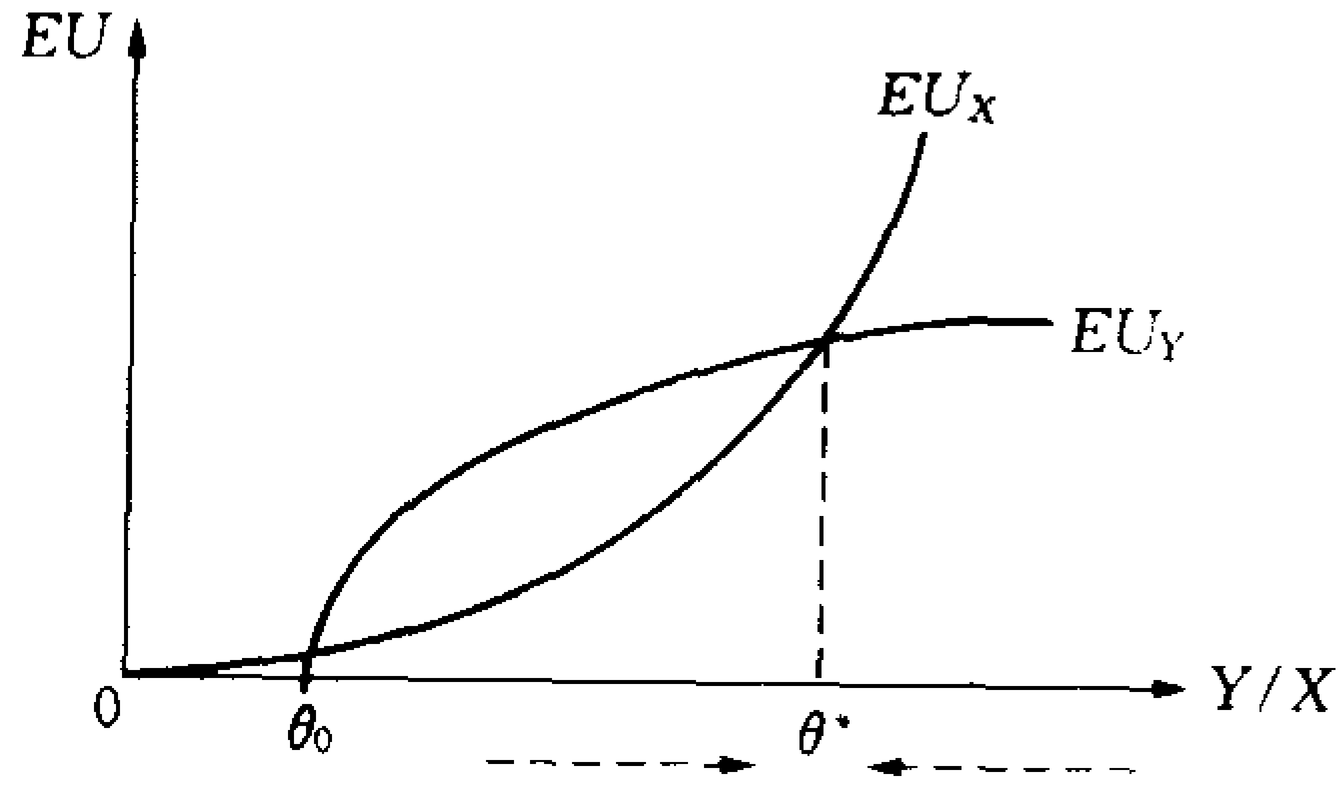

图 3

上图表明，在 $\theta_0 < Y/X < \theta^*$ 区间，利他者的生存适应性 EU_Y 高于利己者的生存适应性 EU_X，利他者的数量趋于增加；在 $\theta_0 > Y/X > \theta^*$ 区间，利己者的生存适应性 EU_X 高于利他者的生存适应性 EU_Y，利己者的数量趋于增加；二者比例收敛于 θ^*。因此，在生物种群及个体行为模式中，利己行为与利他行为将以某种大致固定的比例同时存在。它说明，如果不是孤立地考察利他行为与利己行为的生存适应性，而是在合作及合作剩余的框架中对利他行为与利己行为进行全面、综合的考察，则利他行为完全能够通过整体间的补偿机制体现出相应的进化优势。因此，利他行为与利己行为一样，也是一种“演化稳定策略”。

上图 θ_0 表明，在任何一个生物种群中，由突变、迁徙或其他原因产生的利他者必须超越一个阈值才能得到进化。社会生物学家威尔逊曾经论证，由于亲缘利他的存在，其他利他行为可以从学习和摹仿中滋生。（Wilson，1975）因此，超越这个阈值，在生物长期进化过程中并非是一个不可能的事件。

如上基于 ESS 的模型，不是为了描述利他行为演化均衡的具体过程以及这一过程的动态特征（所以我们忽略了一些技术细节，这方面的研究将另行介绍）。我们的目的旨在说明，主流生物学在这一问题上存在着明显疏漏和严重错误：由于孤立地考察利他行为与利己行为，从而得出前者适应性必然小于后者的结论。我们的模型证明，即使自然选择的基本单位是生物个体，利他行为也可以通过合作剩余条件下的整体补偿机制获得相对的进化优势。

4 结　论

1. 我们认为，主流经济学的偏好模型虽然不排斥利他行为，但主流经济学家长期以来对利他行为的偏见与忽视应该得到纠正。正如桑

塔费学派经济学家所揭示的，在人类漫长的进化历史中，利他行为是人类合作秩序的必然产物。在生存压力特别巨大的环境中，我们的原始祖先不得不进化出一种超越囚徒困境的特殊行为模式，而由所谓的强互惠者实施的利他惩罚就是其中之一。利他主义由于其显而易见的伦理和道德意蕴，往往被人们视为一种“应然”，从而纳入规范性分析的范畴；但就其维持合作剩余不可替代的效率来说，它在事实上仍然体现了一种“实然”，应该纳入实证性分析的范畴。道德与效率、应然与实然之间，不存在无法逾越的鸿沟。正如社会生物学创始人威尔逊所说：“人在过去、现在和将来正是用它来保持人类遗传物质的完整无损，除此之外，道德并没有其他可以证明的最终功能。”(Wilson，1978)

2. 我们认为，主流生物学的个体选择理论虽然得到了现代基因技术和遗传学的支持，但主流生物学家据此得出生物和人类天性自私的结论存在着明显的疏漏和错误。基于 ESS 的演化均衡模型说明，生物适应性是一个综合的评价体系，它不可能被某个单一的事件或关系所决定。如果不是孤立地考察利他行为与利己行为，而是在合作及合作剩余的框架中对生物个体的适应性进行全面评估，即使自然选择的基本单位是生物个体或个体的基因，利他行为也能够通过整体间的补偿机制获得相对的进化优势。利他行为与利己行为都是一种“演化稳定策略”。因此，我们的结论与道金斯、阿莱克什德、杰塞林等主流生物学家的结论大相径庭——自私并不是人类惟一的天性！经过自然选择和进化产生的人类心智与人类行为，不仅与利己心相容，而且也与利他心相容。

3. 我们认为，桑塔费学派经济学家有关利他行为的研究不仅对主流经济学理论体系的完善有重大意义，而且有益于我们更深刻地理解现实的经济活动。最新的社会偏好模型所定义的效用函数包含了对违反公正和合作规范的惩罚愿望，这些模型比经济学传统的自利模型更好地解释了人类的实际行为。由于利他行为不能像自利行为那样从外部获

得物质补偿，因此人类必须进化出一种使行为主体从这些行为本身得到满足的自激励机制。这种机制是对利他偏好内生化的最好说明和描述：大多数人在发现那些违反社会规范的行为未得到惩罚时会感到不舒服，而一旦公正得以建立他们就会感到轻松和满意。也许，这就是上百万年的进化赋予我们人类所特有的道德感和正义感（汪丁丁，2004）。现代社会，包括支撑这一社会的市场交易制度和民主代议制度，在很大程度上依赖人类的这种天性和禀赋。

4. 我们认为，超越囚徒困境中个体理性的局限，谋求合作和合作剩余，可能是我们人类行为、人类心智与人类社会包括人类文化与人类制度共生演化的最终原因。建立一个更完善、更有效率的合作秩序，也许是我们这个物种在生存竞争中的最大优势。在人类漫长的演化历史中，最初的合作秩序是通过自然选择建立的，即自然选择的压力迫使人类进化出有利于合作的偏好，我们把这一阶段称作“自然为人类立法”。随着生产能力的提高，自然施加于人类的选择压力开始减轻，合作秩序不得不通过其他手段来维护，强互惠者个人实施的利他惩罚就是其中之一，我们把这一阶段称作“个人为社会立法”。最后，在近现代社会，工业革命带来的分工使人类合作的规模达到前所未有的程度，合作秩序的维护必须依赖一个建立在民主基础上的现代司法制度，于是我们把这个阶段称作“社会为个人立法”。

5. 最后，我们必须指出，也许和大多数人的认识不同，作为现代经济学理论基础的新古典经济学并非从一开始就排斥对人类天性中的利他主义成分进行分析。恰恰相反，作为新古典经济学创始人的阿尔弗雷德·马歇尔不但没有排斥这种分析，而且认为这种分析是经济学家的最高目标。在《经济学原理》的导言中他曾经指出，“毫无疑问，即使现在，人们也能做出利他的贡献，比他们通常所做的大得多；经济学家的最高目标就是要发现，这种潜在的社会资源如何才能更快地得到发展，如何才能最明智地加以利用。”（Marshall，1890 年）

注释:

[1] 这里指的利他行为不包括亲缘利他和互惠利他，对亲缘利他和互惠利他，经济学与生物学都有比较成功的解释，因为这两种利他行为本质上与自利行为兼容。（叶航，2005）

参考文献:

陈阅增，1997，《普通生物学》，高等教育出版社。

田国强，2005，"现代经济学的基本分析框架与研究方法"，《经济研究》第2期。

汪丁丁，2004，"再谈合作的发生学"，《IT经理世界》第11期。

韦森，2002，《经济学与伦理学》，上海人民出版社。

杨春学，1998，《经济人与社会秩序分析》，上海人民出版社。

杨春学，2001，"利他主义经济学的追求"，《经济研究》第4期。

叶航，2005，"利他行为的经济学解释"，《经济学家》第3期。

张五常，2001，《经济解释》，香港花千树出版有限公司。

Alexander, *The Biology of Moral Systems* (New York: Aldine, 1987).

Arrow, Risk Perception in Psychology and Economics, *Economic Inquiry* 20 (1982).

Becker, *The Economic Approach to Human Behavior* (Chicago: The University of Chicago, 1976).

Bergstrom & Stark, How Altruism Can Prevail in an Evolutionary Environment, *American Economic Review*, Vol. 83, 2, (1993).

Blount, When Social Outcomes Aren't Fair: The Effect of Causal Attributionson Preferences, *Organizational Behavior & Human Decision Processes*, 63.2 (1995).

Bowler, Evolution: *The History of An Idea*, (University of California Press, 1984).

Boyd et al., *Culture and the Evolutionary Process* (Chicago: Univ. of Chicago Press, 1985).

Cavalli-Sforza et al., *Cultural Transmission and Evolution* (Princeton Univ. Press, 1981).

Collard, *Altruism and Economy* (Oxford: Martin Robertson, 1978).

Collard, Economics of Philanthropy: A Comment, *Economic Journal*, 93. (1983).

Darwin, *The Origin of Species* (London: Murray, 1859).

Darwin, *The Descent of Man and Selection in Relation to Sex*, 2nd (London: Murray, 1871).

Darwin & Francis, *The Life and Letters of Charles Darwin*, Including an Autobiographical Chapter, 3 vols, (London: Murray, 1887).

Dawkins, *The Selfish Gene*, (Oxford: Oxford Press, 1976).

Edwards, *Group Selectionism*, (Cambridge: Camb. Univ. Press, 1962).

Fehr et al., The Neural Basis of Altruistic Punishment, *Science*, Vol. 305 (2004).

Fehr & Schmidt, A theory of fairness, competition and cooperation, *Quarterly Journal of Economics*, 114 (1999).

Fehr & Gächter, Cooperation and punishment, *American Economic Review*, 90 (2000).

Falk, Fehr & Fischbacher, *Testing theories of fairness and reciprocity-intentions matter* (Zürich: University of Zürich, 2002).

Gintis & Bowles, The Evolution of Strong Reciprocity: Cooperation in Heterogeneous Populations, *Theor. Popul. Biol.*, 65, 1, (2004).

Gintis, Solving the Puzzle of Prosociality, *Rationality and Society*, 15, 2 (2003).

Gintis, Bowles, Boyd, Fehr, Explaining altruistic behavior in humans, *Evolution and Human Behavior* 24 (2003).

Ghiselin, *The Economy of Nature and the Evolution of Sex* (Berkeley: University of California Press, 1974).

Gossen, *Entwicklung der Gesetze des Menschlichen Verkehrs und der Daraus Fliessenden Regeln Fur Menschliches Handeln* (Berlin: Verlag von R. L. Prager, 1854).

Grusec and Kuczynski, *Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory* (New York: John Wiley & Sons, 1997).

Güth et al., An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization*, 3 (1982).

Hammond, Altruism, *The New Palgarve: A Dictionary of Economics*, Vol. (Macmillan, 1987).

Hirshleifer, Shakespeare vs. Becker on Altruism: the Importance Having the Last Word, *Journal of Economic Literature* (1977).

Hamilton, The Geneticl Evolution of Social Behaviour, *Journal of Theoretical Bio loge*, Vol. 7 (1964).

Jevons, *The Theory of Political Economy* (London: Macmillan and Co., 1871 (1879)).

Lindbeck & Weibull, Strategic Interaction with Altruism: the Economics of Fait Accompli, *Journal of Political Economy*, 96(1977).

Lumsden et al., *Genes, Mind and Culture* (Harvard: Harvard Univ. Press, 1981).

Marshall, *Principles of Economics* (London: The Macmillan Company, 1890 (1938)).

Martin, Evolutionary Fallacies of Nazi Psychiatry: Implications for Current Research, *Perspectives in Biology and Medicine* (2001).

Mas-Colell, Whinston & Green, *Microeconomic Theory* (Oxford University Press, 1995).

Menger, *Pinciples of Economics* (New York: The Free Press, 1871(1950)).

Nakayama, Nassh Equilibria and Pareto Optimal Income Redistribution, *Econometrica* (1981).

Pareto, *Cours d'Economie Politique*, 2 Vols (Lausanne, F. Rouge, 1896).

Simon, *Selections of Simon* (The MIT Press, 1982).

Smith, *An Inquiry Into the Nature and Causes of the Wealth of Nations* (Oxford, At the Clarendon Press, 1776(1880)).

Sugden, On the Economics of Philanthropy, *Econoextmic Journal*, 92(1982).

Trivers, The Evolution of Reciprocal Altruism, *The Quarterly Review of Biology*, Vol. 46(1971).

Tullock, Territorial Boundaries: an Economic View, *American Naturalist*, 121, 3 (1983).

Walras, *Elements of Pure Economics* (Rechard d. Irwinm, Inc., 1874(1950)).

Wiliiams, *Adaptation and Nature Selection: A Critique of Some Current Evolutionary Thought* (Princeton University Press, 1962).

Wilson, *Sociobiology, the New Synthesis* (Harvard: Belknap Press, 1975).

Wilson, *On Human Nature* (Harvard: Harvard University Press, 1978).

Wright, *The Moral Animal, Why we are: The New Science of Evolutionary Psychology* (Pantheon, 1994).

作者中译本序

我们十分高兴中国读者能阅读这些文章。

在这套论文集中，我们运用了演化和行为的方法来研究经济学。这个方法强调个人、社会和经济制度之间的动态互动。这一方法虽然依赖于演化博弈理论、群体生物学以及为行为仿真动态体系提供分析工具的现代数学，但它的基本思想却可以回溯到18世纪末、19世纪初的古典思想家，如亚当·斯密、大卫·休谟、卡尔·马克思的论著中。这个方法在近几年来获得了很大的进展，应该归功于各个学科的学者为之所作的贡献。人类学家、历史学家、计量经济学家和其他许多学科的专家对人类交往行为的深入研究，为这一方法提供了实证基础。最近，实验经济学和行为博弈理论这些新的领域又扩充了该方法的实证基础。与古典思想家一样，这种方法是用一种跨越学科界限的、统一的方法来理解人类行为。

经济学家研究个人行为主要是为了理解这些行为所产生的总效应。例如，他们关心的不是一个具体的个人为什么没有工作，而是关心社会失业率的高低；他们关心的不是一个既定的个人如何交付税款，而是关心自觉交税的人在人群中的分布。知道了一个人的偏好和信仰，以及在特定制度下他们所面临的约束，就可以预测个人的行为。但为了解释总的行为效应，我们就不能仅仅加总所预测到的个人行为，因为每个人的行为都会影响其他人的约束、信仰或偏好。考虑这些反馈效应时必须运用群体行为模型，该模型在群体作为一个整体时把个人的行为与其反馈效应联系起来。

目前，经济学中描述群体行为的方法主要是瓦尔拉斯的一般均衡理论。上世纪中叶，通过阿罗、德布鲁等人的努力，该理论的数学模型得到了极大的完善。这个模型的简单形式不但在经济学，而且在社会科学的许多领域都有广泛的应用，比如在政治选举和婚姻市场上。但是，该模型的缺陷也是显而易见的。中国读者可以在“瓦尔拉斯经济学回顾”一文中看到我们对这一缺陷的讨论。

除了上述一般均衡理论，另一个用来描述群体行为的方法就是那些在变异、遗传和自然选择综合影响下描述生物学系统的动态演化理论。两个方法的相似之处是非常显著的：两个模型都伴随着由竞争所引起的收付转移。因此，达尔文（1809—1882）1838年阅读古典经济学家马尔萨斯（1766—1834）关于人类为生存而奋斗的文章时，想到了“自然选择”这一进化论的核心思想也就不足为奇了。

近几年来，人类学家、生物学家、经济学家运用生物学模型对人类行为的特征进行了研究。这些研究证明，人类行为的许多特征可能是通过基因传递的。在这本文集中，许多文献都涉及了这个问题。这些文献一方面通过修改生物学模型来发展文化演化模型，即考虑人类所特有的能力——凭借拥有的信息，我们可以通过从自己以及他人经历中学到的东西来不断改进我们的生存策略；另一个方面是演化博弈模型，如金迪斯（2000）通过不完全的局部信息来考察我们人类有限的认知能力，从而修改了古典博弈理论。因此，文化演化理论和演化博弈理论分别修改了古典生物学的自然选择模型和古典博弈论的理性模型。第一种情况提高了人类认知能力的假设水平，而第二种情况则减少了其认知能力的假设水平。

读者马上会发现，我们的方法是建立在和许多经济学教科书不同的假设之上的。我们提出的演化和行为方法不是标准的瓦尔拉斯方法（瓦尔拉斯是新古典经济学的伟大创始人之一）或新古典经济学的传统分析方法。我们所说的瓦尔拉斯方法或新古典经济学方法是指：假设个人选择行为是建立在偏好基础上，且行为人对行为后果能够作出具有远见的评价；假设偏好是自利的，而且是外生决定的；假设社会交往只

采取合约化的交流，以及在很多情况下规模收益递增是忽略不计的。这些假设集中体现了瓦尔拉斯方法的特色，以及它在分析上的成功和规范导向。事实上，这个“范式”也是新古典经济学诞生100多年来，我们一直在课堂上教给学生的核心主题。

我们使用术语“演化和行为范式”来指另一个不同于瓦尔拉斯的方法。目前，这个术语还没有形成统一的学派。只是一系列方法的组合，而且很多方法是非常基础的。下面的表格（摘自鲍尔斯，2004）给出了这两个范式的比较。我们希望，它将使中国读者更好地理解这本文集所收录的文章。

“瓦尔拉斯范式”与“演化和行为范式”的比较

	瓦尔拉斯范式	演化和行为范式
社会交往	完全的和可实施的合约，在竞争市场上的双方交往	在非竞争环境下，直接的（非合约的）关系
技　　术	无收益递增的外生生产函数	收益递增的内生技术
更　　新	前向个人在所有知识体系基础上同时更新	后向（基于经验）个人使用局部知识更新
结　　果	建立在个人静态行为上的惟一稳定均衡	建立在群体动态行为上的非稳定的多重均衡
时　　间	静态	动态
变　　异	只涉及风险和保险	演化动力的基本因素
领　　域	经济作为一个自我包含自我约束的实体；偏好和制度外生	经济嵌入一个更大的社会生态体系；偏好和制度共生演化
偏　　好	自利或自涉的偏好；通过结果界定	自利或利他的偏好；通过结果和过程界定
价格和数量	以价格分配资源；行为人不受数量约束	受数量约束；在很大程度上依赖契约订立的时机
方　　法	演绎（从不言自明第一原则中得出）；简约主义和方法论个人主义	归纳（在实验、历史等实证研究上）；非简约主义、在个人或更高次序的实体中选择

在未来的几年中，这个新的演化和行为学的见解是否会成为瓦尔拉斯范式的一个一脉相承的替代还得拭目以待。我们最先通过学习，而后又通过教授瓦尔拉斯范式开始我们的事业生涯。因此，我们深知这一范式的力量，而且不打算从任何方面贬低这个范式。但是，对理解当代社会所面临的基本问题而言，这是远远不够的。这些问题包括古典经济学家很久以前就提出的难题，即如何以我们及我们后代所拥有的知识、劳动和自然环境来改变或规范国家和人们的行为，确保社会财富的长期增长。一个多世纪以前，这些问题曾经激发了马歇尔（1842—1924）和其他新古典范式创始人的兴趣。我们希望，这些问题今天仍然能够激励那些愿意对科学发展作出贡献的人，从而使经济学最后完成这个高尚的使命。

我们谨希望我们的文集会帮助中国的同行也为此贡献自己的力量。

Samuel Bowles, Siena and Santa Fe
Herbert Gintis, Budapest and Northampton
April, 2005

人类合作的起源*

萨缪·鲍尔斯　赫伯特·金迪斯

1 引　　言

人类的合作在自然界中是独一无二的，合作可以扩展到大量相互无关的个体并可采取很多不同的形式。我们对合作的理解是，个体耗费个人成本参加联合活动的行为，其带来的收益要超过引起的费用。比如，这适用于公共品博弈中的捐赠。[1] 尽管在其他物种中这种独一无二的合作形式的缺失可能是演化中的偶然，但更合理的解释是：人类合作是人类某种特殊能力的结果。

基于遗传相关性（亲缘利他主义）和重复互动（如互惠利他主义）对其他物种的合作所作出的一般解释当然也可应用于理解人类的合作。然而，这些机制所包含的能力并不是人类独有的：在很多物种中重复互动和亲缘间互动是普遍存在的。我们并不寻求减小这些类似模式的重要性或建议将它们扩展以说明人类合作的独特方面是不吸引人的。我们宁可寻求一个为人类合作量身定制的新解释，由于其中包含着人类独

* 原文题目为 The Origins of Human Cooperation，Working Paper，周新成译。我们对 Eric Alden Smith 所作的富有帮助的评论以及 Santa Fe Institute、John D. 和 Catherine E. MacArthur Foundation 对这项研究的支持表示感谢。

有的属性（attribute），这种对人类合作的新解释可能不适用于其他物种，或解释力不强。

我们的解释的关键在于人类的认知、语言和自然能力，这些能力使人类可以明确表达人类社会行为的一般规范，使调节这些行为的社会制度得以出现，还使人类有内化规范的心理能力，使族群成员建立在非亲缘特征的种族划分和语言行为的基础之上，这些都有助于减少族群间冲突所引起的高额成本。当然，这并不是要把这些制度和规则看做是先验的。我们必须证明这些制度和规则能同其他人类特征一起以一种合理表现相关环境的方式共同演化。

通过必要的推测，我们把思路限定在三个领域中。首先，我们寻求的对合作形式的解释可由自然观察、历史记录和行为实验来验证。第二，我们要求我们的描述是基于一种将遗传和文化元素相结合的可信的演化动力学，其一致性能够通过正式建模来证明。第三，我们提出的模型的运行必须在能够反映人类生活状态的参数值下对人类的合作作出解释。当模型无法得出分析结果的时候（因为它们太复杂且是高度非线性的），第三个要求使得根据可信参数值的计算机模拟成为必需。

本文内容由以下构成：

- 我们认为，基于亲缘和利他互惠主义的对人类合作的解释是不完全的。
- 我们认为，“强互惠性”是人类个体行为的关键特征，这个特征在很大程度上可以解释人类的合作。
- 我们解释为什么人类族群中基于文化和遗传变异的多层次选择一定在我们的分析中起重要作用。
- 我们发现，一些共同的人类制度创造了条件，在这些条件下多层次选择力量异常强大。这为为什么诸如资源共享和竞争冲突的族群层次的制度，能够和我们称为强互惠的个体行为共同演化提供了一个理由。
- 我们解释了当强互惠者的活动提供了一个很难作假的信号，展现了一般很难被观测到的作为配偶、合作伙伴或对手时的品性时，他们

在演化中是占优的。

我们认为，通过排除来自外面区域的“外来者”所维持的族群边界有益于合作行为的成功演化。同样，这也部分解释了为什么族群成员的特色是决定合作关系范围的因素。

我们认为，人类内化规范以及为了支持合作行为而动用情感的能力削弱了个体利益同群体利益之间的冲突，并且即使在多层次选择和以高发信号成本为特征的合作诱导效应都薄弱的情况下也能支持合作交往。鉴于所收集的资料，我们集中表达一个观点，但并不考虑更多细微和正式的讨论。我们也不考虑其他学者的工作，他们许多人加入了我们这项研究，只想说接下来的文章是近年来同 Ernst Fehr，Simon Gächter，Armin Falk，Urs Fischbacher 和他们的合写者，以及 Robert Boyd，Marcus Feldman，Joe Henrich，Peter Richerson，Eric Alden Smith 等人之间长期合作的结果。我们在这里所表述的观点的意义在我们近年所写的教材中有概述（Gintis，2000a；Bowles，2003）。

2 为什么基于亲缘和互惠利他主义的解释是不完全的

我们不怀疑血缘关系是解释人类合作的重要组成部分，且在其他动物之间也可以这么解释，并且这种亲缘间的合作也许可以逐渐扩展成为非亲缘合作的模板。然而，用这种方法来解释大量没有亲缘关系的个体间的合作是不可信的。

同样的，允许对反社会行为进行报复的重复互动无疑有助于人类或者其他动物间的持续合作。有些观点认为完全自利的人之间的合作演化可以用这种方式来解释。但这是错的。首先，许多关于人类行为有助于合作的实验证据来自于非重复性的互动，或者是来自于重复互动的最后一轮。我们并不认为这些受试者没有发现这是一次性的场景，或

者在实验室进行实验时不能脱离其在真实世界中重复互动的经历。确实，人们很容易区分重复和非重复互动并能相应调整他们的行为，我们对此拥有足够的证据。非实验证据同样明显，因此不能轻易地用对未来互惠的期望来解释日常生活中冲突的通常行为。

第二，早期人类的环境可能使得重复报复机制对合作的支持无效。游猎成员可以通过加入其他族群而逃避报复。而且，在许多人类演化的关键情形中，当冲突或逆境导致族群面临解散时，不太可能发生重复互动。

第三，在有大量人员互动的情境下，能够用重复互动和报复来解释为什么自涉者会合作的条件无法得到满足。经常用来说明自涉个体间的重复互动能够支持表面上的他涉（other-regarding）行为的著名定理是“无名氏定理”，但它不能很可信地从两人扩展到 n（n 值较大）个人组成的族群。在这方面，两人之间和 n 人之间互动的关键差异在于：(a) 偶然和故意的背叛的数量将随着 n 的增大而增大，而这个“颤抖”将显著增大对背叛者进行惩罚的成本；(b) 族群中很大一部分异质行为者有足够的远见预测到 n 上升时合作利益会急剧下降；(c) 协调与激励机制要求自涉族群成员对背叛者进行惩罚，在 n 上升的时候这变得非常复杂和难控制。[2] 尽管许多重要的人类互动是双人的（比如，商品的双向交易），但许多合作的重要例子（比如，通过共同保险减低风险，信息共享，维持有利于族群的社会规范，族群防卫）都是大规模群体间的互动。就这些事情而言，无名氏定理不能为合作是普遍、长久而不是稀少、短暂的提供理由。

3 利他主义的心理和行为方面：趋社会情感和强互惠性

趋社会情感是一种导致行为者从事我们前面所定义的合作行为的生

理和心理反映。一些趋社会情感，包括羞耻、负罪感、同情，以及对社会制裁的敏感性，导致行动者承担建设性的社会互动行为；其他的，如对背叛规范者进行惩罚的愿望，当趋社会情感不能在社会族群的一部分人中诱发足够的合作行为时有助于减少搭便车行为（Frank，1987；Hirshleifer，1987）。

如果没有趋社会情感，不管怎样加强契约制度、政府法律的强制力和提高声誉，我们都会成为反社会的人（sociopaths），而且人类社会将不会存在。除了体验羞耻、负罪、同情和懊悔的能力严重受到削弱或缺失外，反社会的人没有任何智力障碍。他们在美国男性中占3%—4%（Mealey，1995），却占美国囚犯数量的大约33%，以及惯犯人口的33%到80%。

趋社会情感是产生善意和关爱行为的源泉，这些行为丰富了我们的日常生活并使我们在陌生人中生活、工作、购物和旅游时感到舒适和愉快。此外，代议制政府、公民自由、妇女的权利、对少数民族的尊重，这些没有人类尊严就不会在现代社会中存在的关键制度都是在人们参与集体行动的过程中产生的，不仅让他们追求个人目标，同时也关怀整个人类。我们的自由和舒适等等都是基于过去几代人的情感积淀。

尽管趋社会情感可以解释人类合作重要形式的证据确凿，但还没有普遍认可的关于情感如何与认知过程相结合来影响行为的模型。现在对如何最好地描述支持合作行为的趋社会情感也没有一个统一的意见，尽管我们（Bowles and Gintis，2002）已经在这个方向上作了尝试。但是，对于自涉偏好不能解释很多善行的断言是不存在争议的，这些行为包括为什么人们投票，为什么人们会匿名行善，为什么人们会在战斗中牺牲自己。在涉及社会生活的这些领域时，证据指向了我们所说的强互惠性行为。一个带着合作倾向进入一个新社会环境的强互惠者，倾向于通过维持或提高他的合作水平来对其他人的合作行为作出回应，并对他人的“搭便车”行为进行报复，即使这会让他花费成本，甚至不能理性预期这种报复能否会在将来给个人带来收益。这个强互惠者既是有条件的利

他合作者也是一个有条件的利他惩罚者，他的行为在付出个人成本的时候会给族群其他成员带来收益。我们称其为“强互惠”，以区别于其他如互惠利他主义、间接互惠以及由重复互动或正相配（positive assortation）所维持的安排个体自涉行为互动的“弱互惠”（见 Fehr et al.，2002）。

4 多层次选择

群体成员内部之间交往远比同外部人交往的多得多，我们早就知道演化过程可以分解成群间和群内选择效应（Price，1970）。在特征复制率依赖于族群成分以及族群成分存在差异的情况下，族群选择对演化变化的速度和方向有影响。然而，直到最近，大多数将群体和个体选择共同影响的演化过程模型化的学者得到的结论是前者不能抵消后者，除非是在特殊的环境因素（小族群，有限移民）升高和维持相对于群内差异的群间差异的地方。

因而，普遍认为族群选择模型不能解释利于群体但对个体是高成本的行为的成功演化。相对于其他动物而言，基于遗传和文化变异上的族群选择对于人类具有更重大的意义。人类特有的增强族群选择相关性的特征使我们具有压制群内表型差异（比方说，通过资源共享、共同保险和一致决策），盲从的文化传递，民族优越感（以支持群内调和并维持族群边界）以及群间频繁冲突的能力。

在金迪斯（2000b）的论文中，我们提出了一个解析模型，该模型显示，在可行的条件下，强互惠性可以通过族群选择在互惠的利他主义中产生。这篇论文中，合作被模拟为在一般的条件下，行为者对将来能从其他成员那里得到回报有充分认识的 n-人公共品重复博弈，如同无名氏定理中所断言的那样，合作通过触发策略得以持续。但是当族群面临灭绝或者解散的威胁时，比如说遇到战争、瘟疫或者饥荒，为了生存，合作将变得更为必须。然而，当族群受到威胁时，一个人对族

群的贡献在将来得到回报的可能性急剧下降，群体解散的可能性上升，从而合作的动机也会不复存在。因而，恰恰当一个族群最需要趋社会行为时，基于互惠利他主义的合作将崩溃。在我们这个物种演化的历史上，这种关键时期是很常见的。少量不考虑未来的回报而对背叛者施以惩罚的强互惠者，能够显著提高人类族群的生存机会。而且，人类是群居且个体具有以低成本对受罚者施加严厉处罚的能力的惟一物种，这是人类高超的制造工具和狩猎能力的结果。确实，同其他灵长类作比较，即便是最强壮的人在睡觉时也能被最弱的人轻易杀死。使用 Price 方程可以简单证明，在这些条件下，强互惠者能够侵入一个自涉类型的人群并保持均衡。

我们同 Boyd 和 Richerson（Boyd et al.，2003）的合作研究结果显示，通过基于行为人的模拟，对某些合作行为（显著地惩罚那些违反合作规范者而言），基于文化传递特征的族群选择即便是在非常大的族群和存在大比例移民的族群中仍然具有决定意义。产生这个令人惊讶的结果的原因在于，如果族群的大多数成员坚持规则，事先确定的对违反者进行惩罚所引发的成本是非常小的，因为背叛行为并不常见。因而，即使反合作行为的群内选择存在，它在合作均衡的附近也是非常弱的。这点说明了不同的族群长期存在很大差异：一些族群内几乎完全是合作的行为者，都事先决定合作并对不合作者加以惩罚，而另一些族群几乎完全由自涉的行为者组成。胜利的族群吸收失败者并随后分化导致了其他的群间变化。

这些模型中一个特别吸引人的特点是，它们预测了一个异质的均衡，如同在实验文献中发现的那样，自涉者和强互惠者都包含在这个均衡中（Fehr and Gächter，2002）。

5　制度和行为的共同演化

如果族群选择是对个体合作行为的成功演化的部分解释，那么可以

认为增强族群选择压力的族群特征（比如，相对较小的族群规模，有限的移民，频繁的群间冲突）是同合作行为共同演化的。这种族群特征和个体行为可能存在协同效果。合作部分基于人类某种特殊的能力，它可以构建某种制度环境，以限制群内竞争、减少群内表型变异，同时提升群与群竞争的重要性，并允许代价昂贵但利于群内的个体行为通过群间选择过程同那些支持性的环境共同演化。

这种对群内竞争进行压制可能会严重影响演化动力的观点可以在群居昆虫或其他物种中得到广泛认可。Alexander（1979）、Boehm（1982）和 Eibl-Eibesfeldt（1982）首先将这种推理运用到人类社会，探究文化传递的惯例对减少群内表型变异的作用。这些作法中有均等制度，如在非亲属间分享资源，以及减少群内繁殖适存度或物质福利的差异等，它们可以减少在福利缺乏时群内成员所获得的显著差异。因此，尽管存在对群内其他成员慷慨的优秀猎人可能比其他猎人拥有较高的适存度和较好的营养（作为消费比较平稳的结果）的事实，但这并不表明缺乏均等，除非这些惯例同样会导致不太成功的猎人的适存度和营养的恶化（这似乎不太可能）。

通过减少个体成功方面的群内差异，这些惯例可能会削弱那些压制对族群有利但对个人有成本消耗的举措的群内遗传或文化选择的操作，因而族群可以在群间竞争时得到这些利益。族群层次的制度因此创造了能够告知生物演化和文化改变过程明确方向与速度的环境。因此，减少了不利于群内利益的个体特征的选择压力和具有这些特征的人可以有效地减少族群灭绝可能性这些事实可以解释减少群内表型差异的社会制度的成功演化。

我们沿着这些新的特征建立一个演化动力学模型。这些通过文化、遗传传递的个人行为以及通过文化传递的族群层次的制度特征受制于选择，并且群间竞争对族群层次的选择起了决定作用（Bowles，2001；Bowles et al.，2003）。我们可以看到，群间冲突可以解释下面两个成功演化：（a）人类社会的利他形式转向非亲缘，（b）族群层次

上的制度结构，例如资源分享，在人类漫长历史进程中重复出现并扩散到各种生态系统。如果对群外成员施加惩罚的成本足够高而族群层次上的制度能够限制这些行为所带来的成本，减弱反对这些行为的群内选择，那么有利于群内利益的行为就可能得以演化。

我们的模拟显示，如果族群层次上贯彻资源共享或者族群成员非随机配对的制度能够演化，那么利于族群的个体特征就能和这些制度共同演化，即便后者会对采纳它们的族群施加高额成本。这些结论在下面的说明中保持不变：人群中的个体合作行为和社会制度在开始是缺失的。然而，在缺乏族群层次的制度时，只有当群间冲突非常频繁、族群较小且移民比例较低时，有利于族群的特征才能演化。因此，存在于90 000年前的解剖意义上的现代人在相关环境下合作行为的成功演化也许是人类特有的建构社会制度的能力的结果。

6 作为品质信号的强互惠

因为增加了个体交配和共同建巢的机会，合作行为在演化过程中将可能受惠。如果其他人把有价值信息的共享和保卫族群引致的危险当作个人特质的诚实信号（这些作为配偶和政治同盟的特质不然是无法观察到的），那这就可以成为一个例子。许多关于昂贵的信号和人类演化的文献解释了诸如好的猎人将他们的猎物贡献给其他人的行为，同样的推理可以应用到合作行为。合作行为因此能够产生以这种信号方式结成的联盟，这个结盟产生的增强适存度和物质成功的结果能够解释构成这个信号的合作行为的增加。和 Eric Alden Smith 一起（Gintis et al.，2001），我们将这个过程模型化为多人公共品博弈。这个博弈不包含重复或协调互动，因此如果没有明显的信号收益，非合作将成为占优策略。我们可以注意到，通过为群体其他人提供一项公共品发出关于品质的诚实信号传递能够稳

定地演化，并且如果拥有确定可行的条件，可以在最初稀少的人群中扩散。因此，与我们所说的强互惠性相一致的行为可以通过这种方法演化。

然而，单单信号均衡并不需要这个信号赋予族群其他成员利益。反社会的行为可以执行同样的功能：痛打邻居以显示威力与在群体防御中表现勇敢是差不多的。如果发出信号是一种有利于群体行为的解释，本模型的逻辑必须通过证明有利于群体的信号要优于反社会的信号而得到补充。我们通过指出所提供公共利益的水平与发信号者提供给那些对信号作出反应的人的利益是正相关的来证明这一点。例如，防卫族群的人将比那些敲打邻居的人更有可能给予他的同伴或盟友利益（比方说，保护）。有利于族群的信号比反社会的信号更能吸引大众。最后，相互竞争族群的族群选择将青睐那些使族群受益的信号均衡而不是那些非信号均衡或反社会信号均衡。

如同这个最后的原因所提出的，发信号的族群选择对合作的作用是促进而不是简单的累加。族群选择解释了为什么发信号是趋社会性的，信号理论提供了为什么族群受益的行为在群内动态过程中可能是演化稳定的，从而有助于行为的族间差异并因此增强了族群选择力。

7 狭隘主义（parochialism）和互惠

个体表现合作的倾向通常依赖于对他们交往的人的认同："内部人"优于"外部人"。在上面的模型中，内部人—外部人的差异充当着关键的角色。在我们的族群选择模型中，合作行为给予同族群成员以利益并允许高合作的族群从族间冲突中胜出。某些有利于族内成员的特定行为对其他族群成员是有代价的，甚至可能是致命的。另外，如我们所看到的，维持族群的边界以限制移民数量以及族群间冲突的频率，这为提升群内合作的族群选择提供了持续的动力。于是，似乎可

以说群内的合作与对“外部人”的敌意是共同演进的。

群内的偏爱经常受到物质上、语言上和其他标志内部人的特征的文化优势的支持，并对外部人施以排他性行动，我们把这个行为称为狭隘主义。我们将狭隘主义模型化为具有可互动的归属特征并排除相反特征的一个过滤器（Bowles and Gintis，2000）。群内成员通过两种方式从采用狭隘过滤器中受益：均衡的族群规模的缩小以及族群成员文化差异的降低，这都将增强互相监督和建立声誉的有效性，支持群内高水平合作。更多的小族群在形成规模经济、通过交换获利，以及从更多不同的成员中获得可能的集体认知收益方面居于领先地位。观察到的狭隘程度取决于这些利益同排他性行为的成本之间的平衡。由于所有这些都随着环境和技术的改变而演化，我们关于“优化的狭隘”的分析也许可以提供一个关于合作与群外敌对共同演化的模型，尽管我们没有尝试这一宏大的计划。

8　趋社会情感：模型和实验证据

如同我们前面论证的，忠实于社会规范不仅是出于追求自利的认知考虑，同时也是出于情感。羞耻、同情和其他内心活动在维系合作关系中起了很重要的作用。令人困惑的是趋社会情感乍看起来是一种利他的行为，能够给别人带来利益而自己花费成本。在动态收益单调的情况下，自涉的特征一般会增加频率，因此趋社会性将萎缩。

痛苦是一种趋社会的情感。羞耻是一种社会情感：当一个人因为违背一种社会价值或没有遵守一种行为规范时，他会因被他所处的社会群体的其他人贬低而感到痛苦。

羞耻能够实现类似痛苦的那种效果吗？如果被社会贬低具有适存度成本，而且羞耻的数量与福利成本的水平密切相关，那么答案就是肯定的。像痛苦一样，羞耻使行为者去弥补导致这种刺激的情

况并在将来避免类似的情况。如同痛苦那样，羞耻用一种简单的信息取代复杂的优化过程：不管你在做什么，可能的话停下来，且不要再做。当然，如果好处足够大，个人也可以克服这种不愉快的羞耻感，但是在平均水平上，这种情感仍然能够减少导致羞耻的社会行为的发生频率。

既然羞耻是一种演化选择而且具有成本，平均来讲它应该给予那些体验过羞涩的人一些选择优势。两类选择优势在这里起作用。首先，羞耻可以增强一个具有不完全信息（比如，一个特定的反社会行为是如何减少适存度的）、有限或不完善的信息处理能力以及倾向于低估未来再行动的成本与收益的行为人的适存度。或许在没有羞耻的情况下，所有的三种条件将共同导致对社会非难的次优反应，而羞耻使得我们更接近于最优。羞耻有警示我们将来的消极结果的作用是以所组成的社会会对规则违反者施加惩罚为前提的。羞耻可能是与促使对反社会行为进行惩罚的情感共同演化的（我们模型中的互惠动机）。

那些有羞耻感的人的第二种选择优势通过族群竞争效应而体现。在羞耻感非常普遍的地方，对反社会行为的惩罚将非常有效，其结果是惩罚很少被采用。于是，普遍有羞耻感的族群将以有限的成本维持一个较高的合作水平并通过群体间选择得到扩张。羞耻于是成为一种节省成本的群内惩罚的方法。

羞耻可以在实验室内进行研究。鲍尔斯和金迪斯（2002）研究了一项行动者最大化包含五项不同动机的效用函数的公共品博弈，这些动机有个人的物质收益、对别人收益的评价、对利他性和互惠程度的依赖和由于没有能够对集体努力作出同等贡献带来的负罪感或羞耻感。如果被别人惩罚过的参与者以比一个物质收益最大化行为者更为合作的行为来回应，那么羞耻感就是存在的。我们提供间接的实验证据来说明这种情感在公共品博弈中起着某种作用。但是，博弈中情感作用的直接证据仍然稀缺。

9 规范的内化

内在规范是部分通过内在制裁来强制实施的一种行为模式。当人们自己评估某些行为以及/或不论这些行为对个人适存度和/或可觉察到的福利的影响时，他们在遵循内在的规范。将规则内化的能力在人类中是很常见的。虽然规范内化在社会学文献（社会化理论）中有广泛的研究，但在该领域之外几乎完全被忽略。

社会化模型因为提出人们采取的规范独立于他们的认知收益（perceived pay off）而遭到严厉批评。事实上，人们并不总是盲目遵循那些社会规范，有时他们将顺从作为一种策略选择（Gintis，1975）。在社会学文献中提出的“过度社会化”（oversocialized）模型中，个体能通过增加一个反映行为者从低收益转向高收益策略的表型复制过程而获得平衡（Gintis，2003b）。

所有成功的文化都鼓励增强个人适存度的内在规范，如面向未来、好的个人卫生、积极的工作习惯以及控制情感。文化同样普遍地促进使个体从属于群体福利的利他主义的规范，如勇敢、诚实、公平、积极合作以及对他人痛苦的同情。

给定大多数文化促进合作行为，并且如果我们接受社会学关于个体内在规范是通过父母和其他有影响力的长辈来传递的观点，解释人类的合作将变得简单。只要社会的一小部分人将合作的规范内化并惩罚搭便车者和其他违反规范者，高度的合作就能够长时间得到保持。这样我们有两个谜题：我们为什么要将规范内化？为什么文化能促进合作行为？

我们提供一个演化模型，在这个模型中内化能力得到发展，因为这种能力可以在一个社会行为非常复杂和多面以至于不能通过个人理性评估获得成功评判的世界中增强个体适存度（Gintis，2003a）。内化把

规范从个人实现福利最大化的手段变为终极目的。不难显示，如果一项内化的规范能提高个体适存度，为了可行的社会化模式，内化规范的等位基因是稳定演化的。

我们（Gintis，2003a）使用这一框架来模拟赫伯特·西蒙对利他主义的解释。西蒙认为利他性规范可以搭内化规范一般倾向的“便车”增强适存度。但是，西蒙没能提供这个过程的正式模型，而且他的观点完全被忽略了。本章指出西蒙的见解能够解析地进行建模，而且在可行的条件下是有效的。一场直截了当的关于基因—文化共同演化的辩论就解释了为什么降低适存度的内在规范是趋社会的而不是有害社会的：有着趋社会内在规范的族群将胜过那些有着反社会或中立的内在规范的族群。

10 结　　论

我们对人类合作的说明中贯穿着两个主题：（a）人类演化中族群的重要性和多层选择的力量。（b）潜在动态的基因—文化共同演化。我们以对被我们认为是错误的两种方法进行评论作为结束，它们是把自利以同义反复的形式拓展到演化基本法则的地位以及将文化当作基因和环境互动的一种附带表达。

如同托克维尔笔下的“美国人”，生物学和社会科学的一个有名的传统已经开始寻求用“正确理解自利的原则”来解释合作行为。从 J. B. S. Haldane “他将冒着生命危险去救八个溺水的兄弟姐妹”的双关语到现代博弈论中的无名氏定理，这种传统已经阐明了血缘关系、重复博弈，以及群体成员间社会互动的其他方面可以赠予那些从事无私行为的人适当的好处。关键在于，如果注意到通过遗传和文化传递的特征选择的差异复制是收益单调的，那只有平均高收益的特征才能成功地演化。如果自利行为被定义为具有平均高收益特征的行为，那么自利原则就成为演化的基本法则。

一些著名的演化生物学家正好坚持这一点。例如道金斯（1989）在《自私的基因》的前四页中写道："在一个成功的基因中可望发现的显著特征是无情的自私性。这种基因的自私将产生个人的自私行为……让我们试着传授慷慨和利他主义，因为我们生来自私。"[3] 类似地，从人类行为的演化分析中抽象出哲学含义，Richard Alexander（1987）说："只有社会被看作是个体追求他们自己个人利益的集合时，我们才能理解伦理、道德、人类行为以及人类心灵。我相信，人们通常服从于他们所能觉察到的个人利益才是人类行为的最基本准则。"

如托克维尔一样，我们反对将自利概念作同义反复地扩展。我们不关心这个方法中假定的基于适存度或其他收益单调的动态过程。毫无疑问，人群中适存度较低的特征在可行的动态演化中将遭受阻碍：甚至文化的演化也会偏向于导致个体物质上成功的行为。而且，我们注意到对"自利"这个概念的曲解。在达尔文的词语中，诸如"随时准备在危险的时候警告对方、相互帮助和彼此保卫"都被认为是"自私"的同义词。像达尔文（1871／1893）认为的，普遍存在这些行为的部落将"扩张并战胜其他部落"。我们避开"自私"和"自利"两个术语以避免混淆，根据对个体的成本和对群体成员的利益来定义合作行为。我们的模型和模拟显示，在合理情况下这些行为将增加，这是人类群体的族群结构和合作者普遍存在的族群成功的结果。

转到我们要讨论的第二个要点。我们注意到，各种不同的作者，如马克思和一些当代的社会生物学家，在他们的论述中普遍（即便很少直接提出）将文化缩影为遗传和自然环境交互作用的一种结果。如同自利原则，那些认为自然环境和基因交互作用影响文化演化的假说引发了众多的洞见。然而，同样正确的是，文化会影响到某些自然和社会环境，在这个环境里基因传递的行为特征的相对适存度是已经决定了的。Cavalli-Sforza 和 Feldman（1981），Boyd 以及 Richerson（1985），Durham（1991）和其他一些人提供了一些文化影响遗传演化的引人注目的例子。我们自己的通过基因遗传个体行为和通过文化传

递群体层面上的制度这两者共同演化的模型只是这个过程中许多模型中的一部分。例如，在我们的一个模型中，我们看到通过文化传递的习俗（资源共享）对于由自然选择所控制的通过基因遗传的利他特征的演化具有重要作用。把人类文化，尤其是它们所支撑的制度结构表达为一种生态位构架，也就是一种特殊的、影响遗传演化的环境，这会是很有帮助的（Laland et al.，2000；Bowles，2000）。

对人类合作的起源的解释所作的挑战引导我们对以游猎为生的人类或者是无国家的简单人类社会及生活环境进行研究，他们可以无可争议地构成解剖学意义上的现代人并构成历史上的人类社会。同样的探索使得非合作博弈理论（假设不存在可实施的参与前协议）成为一种基本工具。但是，如同几个作者所指出的，大多数当代合作是以多边对等交往和第三方强制的实施为基础，并由现代国家来实现的。或许，避免根据我们对晚更新世的合作的起源的思考来对 21 世纪的合作下冒然的结论是谦逊甚至是明智的。

注释：

[1] 合作的这种定义排除了互惠交往（互利共生，进化理论对此的解释更为简洁）、利他主义的非生产性形式（这时得到的利益不会超过利他主义者的成本），以及缺乏共同收益的联合活动这些我们渴望解释的行为特征。

[2] 一个著名的定理显示，一大群行为者中的重复能够支持有效率的合作均衡（Fudenberg and Maskin，1990），这种均衡高度要求族群成员永生，甚至不能近似地应用于那些拥有最乐观地长寿假设和未来安排条件的人类族群。

[3] 注意这最后一句中的倾向以鉴别具有一般意义上高收益的自利，但使用该术语在日常生活中的含义是完全没有根据的。

参考文献：

Alexander，R. D.，*Biology and Human Affairs* （Seattle：Univ. of Washington Press，1979）.

Alexander，R. D.，*The Biology of Moral Systems* （Hawthorne，NY：de Gruyter，1987）.

Boehm，C.，The evolutionary development of morality as an effect of dominance behavior and conflict interference，*J. Soc. Biol. Struct*. 5 （1982）：413—421.

Bowles，S.，Economic institutions as ecological niches. *Behav. Brain Sci*. 23 （2000）.

Bowles，S.，Individual interactions，group conflicts，and the evolution of preferences. In：Social Dynamics，ed. S. N. Durlauf and H. P. Young，pp. 155—190 （Cambridge，MA：MIT Press，2001）.

Bowles，S.，*Microeconomics：Behavior，Institutions，and Evolution*，Princeton

(NJ: Princeton Univ. Press, 2003).

Bowles, S., and H. Gintis, Persistent parochialism: The dynamics of trust and exclusion in networks, Santa Fe Institute Working Paper 00-03-017, Santa Fe, NM: Santa Fe Institute.

Bowles, S., and H. Gintis, Prosocial emotions, Santa Fe Institute Working Paper 02-07-028, Santa Fe, NM: Santa Fe Institute.

Bowles, S., J.-K. Choi, and A. Hopfensitz, The coevolution of individual behaviors and group level institutions, *J. Theor. Biol.*, in press.

Boyd, R., H. Gintis, S. Bowles, and P. J. Richerson, Evolution of altruistic punishment, *Proc. Natl. Acad. Sci. USA* 100 (2003): 3531—3535.

Boyd, R., and P. J. Richerson, *Culture and the Evolutionary Process* (Chicago: Univ. of Chicago Press, 1985).

Caporael, L., R. Dawes, J. Orbell, and J. C. van de Kragt, Selfishness examined: Cooperation in the absence of egoistic incentives, *Behav. Brain Sci.* 12 (1989): 683—738.

Cavalli-Sforza, L. L., and M. W. Feldman, *Cultural Transmission and Evolution* (Princeton, NJ: Princeton Univ. Press, 1981).

Darwin, C. 1871 (1973), *The Descent of Man* New York: Appleton Press.

Dawkins, R., *The Selfish Gene*, 2d ed (Oxford: Oxford Univ. Press, 1973).

Durham, W. H., *Coevolution: Genes, Culture, and Human Diversity* (Stanford: Stanford Univ. Press, 1991).

Eibl-Eibesfeldt, I., Warfare, man's indoctrinability and group selection, *J. Comp. Ethnol.* 60 (1982): 177—198.

Fehr, E., and S. Gächter, Altruistic punishment in humans. *Nature* 415 (2002): 137—140.

Fehr, E., U. Fischbacher, and S. Gächter, Strong reciprocity, human cooperation and the enforcement of social norms, *Nature* 13 (2002): 1—25.

Frank, R. H., If Homo economicus could choose his own utility function, would he want one with a conscience? *Am. Econ. Rev*. 77 (1987): 593—604.

Fudenberg, D., and E. Maskin, Evolution and cooperation in noisy repeated games. *Am. Econ. Rev.* 80 (1990): 275—279.

Gintis, H., Welfare economics and individual development: Areply to Talcott Parsons. *Qtly. J. Econ.* 89 (1975): 291—302.

Gintis, H., *Game Theory Evolving* (Princeton, NJ: Princeton Univ. Press, 2000a).

Gintis, H., Strong reciprocity and human sociality, *J. Theor. Biol.* 206 (2000b): 169—179.

Gintis, H., The hitchhiker's guide to altruism: Gene-culture coevolution and the internalization of norms, *J. Theor. Biol.* 200 (2003a): 407—418.

Gintis, H., Solving the puzzle of human prosociality, *Ration. Soc*, 15 (2003b).

Gintis, H., E. A. Smith, and S. Bowles, Costly signaling and cooperation, *J. Theor. Biol.* 213 (2001): 103—119.

Hirshleifer, J., Economics from a biological viewpoint, In: Organizational Economics, ed. J. B. Barney and W. G. Ouchi, pp. 319—371 (San Francisco: Jossey-Bass, 1987).

Laland, K., F. J. Olding-Smee, and M. Feldman, Group selection: Aniche construction perspective, *J. Consc. St.* 7 (2000): 221—224.

Mealey, L., The sociobiology of sociopathy, *Behav. Brain Sci.* 18 (1995): 523—541.

Price, G. R., Selection and covariance. *Nature* 227 (1970): 520—521.

Simon, H., Amechanism for social selection and successful altruism, *Science* 250 (1990): 1665—1668.

社会资本和共同体治理*

萨缪·鲍尔斯　赫伯特·金迪斯

1 引　言

社会资本通常涉及信任、关心同事、自觉遵守共同体规范以及自发惩戒违规者。从亚里士多德到托马斯·阿奎那到埃德蒙·伯克（Edmund Burke），这些古典思想家都将此类行为看做是有效治理的必要因素。然而18世纪后期的政治理论家和宪政思想家（constitutional thinkers）则把经济人当作分析的出发点，并且在此基础上强调其他必要因素（desiderata），特别是竞争的市场、明确界定的产权和有效且具备善良意志的政府，从而使好的游戏规则代替了优秀市民（good citizens）成为有效治理的必要条件（sine qua non）。

19世纪到20世纪初出现的两大阵营界定了治理所需的很多制度和政策条件，争论的一方拥护自由放任（laissez faire）而另一方则支持全面的政府干预。实践型的人（Practically-minded people）出于良心或

* 原文题目为 Social Capital and Community Governance，发表于 *Economic Journal* 112 (2002)：419—436，熊艳艳译。感谢 Kate Baird，Michael Carter，Jeff Carpenter，Christina Fong，Yujiro Hayami 和 Elisabeth Wood 的帮助和评论，以及 John. D 和 Catherine T. MacArthur 基金会的财政资助。

是受到选举的约束，很少以武断的态度去寻求解决社会问题的办法，从不简单狭隘地接受争论中任何一方的观点。不过这场争论活跃了学术界的气氛，20 世纪中期甚至后期比较经济制度的课本中就包含了相关的内容。争论双方共同的隐含假设是市场或政府都能充分控制经济过程；除了市场和政府别无其他的资源配置方式，对市场和政府进行混合是不可能的。但是当时争论的主要观点、那些维护自发秩序或是社会计划的夸张言辞，在今天看来已经过时了。在 20 世纪即将结束之时，争论双方都不再执着于自己的立场，越来越多的人逐渐认识到同时存在着市场失灵和政府失灵。社会资本逐渐引起人们的重视并不是因为它自身的优点，而是由于市场和政府都存在缺陷。

主张政府干预的一方重视社会资本是因为它肯定了在解决社会问题的过程中信任、慷慨和集体行动这些因素的重要性，从而反对这样一个观点：明确界定的产权和竞争的市场可以成功地驱使自私动机实现公共目标，至于公民美德则并非必要。自由放任的支持者对社会资本着迷则是因为它表明，在市场失灵（如提供地方公共品和各种保险）的地方，住宅区、父母教师协会和保龄球社团等组织而非政府就可插手解决这些问题。

如果不是政府的能力和责任的局限性在官僚制度的自负和不切实际的五年计划中得到了无误的证明，美国的自由主义者、社会民主主义者和市场社会主义者也许不会加入这场争论。如果保守派理想化的制度能够运作得好一些，那么他们的渴求也许就会少一些。但是 20 世纪早期出现的大萧条、增长的环境问题以及不平等现象的加剧，使得教科书中理想化的资本主义失去了光泽。这些幻想的破灭为社会资本的引入建立了基础。

10 年前，一些在其他方面持怀疑论的学者和疲惫的政策制定者指出意大利托斯卡纳区合唱队的社团和该地的有效治理存在高度相关性，还对一个公民意识渐缺（bowl alone）的国家提出预警，并引用托克维尔对美国作为一个社区国家的评论，这使他们的朋友感到非常吃惊。乔治·布什（老的那个）总统一直极力主张美国人民进行政府转向，建立起一个富有活力的市民社会。前第一夫人希拉里也提出“应该让乡

村去抚养孩子”，世界银行还专门为这个题目建立了一个网站。

社会资本的兴起反映了政界和学术界日益重视真实的人的价值观，它并不仅是经济人的一个经验上可信的效用方程，它强调人们在日常生活中如何与家庭成员、邻居、同事交往，而不仅仅是局限在作为买者、卖者和市民的身份上。社会资本概念的兴起也表明了充满意识形态色彩的“计划 vs. 市场”争论的结束。

如同伏尔泰的上帝一样，“如果社会资本原本不存在，它也可以被创造出来。”这是一个不错的想法，但并不是一个好的说法。资本指可以拥有的物品，即使是孤立社会中的鲁宾逊·克鲁索也拥有一把斧头和一张渔网。对比资本的概念，社会资本的意义在于描述人与人之间的关系。“共同体”很好地反映了有效治理的方方面面，因为它关注于群体在做什么，而不是拥有什么，这解释了社会资本的流行。提到共同体，我们是指这样一群人，他们直接地、频繁地、多方面地互相接触和互相影响。一起工作的人通常构成这种意义上的共同体，像住宅区、一群朋友、专职人员和商业网络、帮派、运动联盟都是共同体。这些因素说明，人与人之间的联系而非感情才是构成共同体的最主要特征。无论一个人生于一个共同体，抑或是主动进入一个共同体，在通常情况下，迁徙到另一个共同体都要付出巨大的代价。

共同体这个概念使我们明白，要理解信任、合作、慷慨以及其他社会资本研究所涉及的概念时，我们必须研究社会交往的结构，弄明白同一个人为何在不同场合的社会交往中表现出不同的社会资本水平与类型。这条从社会结构入手研究社会资本的路径可能会和 Glaeser, Laibson 和 Sacerdote（2002）提出的基于个体的社会资本模型形成对比。

在接下来的文章中，我们将提出一个新的“共同体治理”框架。我们先举一些例子，给出一个简单的模型和一些实验结果，证明基本行为假设的合理性。与 Durlauf（即将发表）强调的社会计量问题不同，我们怀疑那些常用的指标是否能精确预测我们的实际行为。例如，Glaeser，Laibson，Scheinkman 和 Soutter（2000）发现，由福山

(1995) 和其他人提出并被广泛接受的关于衡量信任的那些标准问题，不管是受试者在现金激励的实验中的反应，或者是他在现实生活中的行为（如是否愿意借款给别人），都没有提供任何新信息。接着我们回到共同体治理的本土性问题，以及那些和我们一样确信政策设计必须认识并且致力于加强市场、政府以及共同体之间互补性的人所提出的挑战。[1] 最后，我们将以对共同体未来重要性的一些思考结束本文。

2 共同体治理

共同体是有效治理的重要组成部分，因为它可以处理一些个人无法独立解决，市场和政府也很难解决的问题。

例如 Felton Earls, Robert Sampson 和 Steven Raudenbush (1997) 研究了芝加哥的一些住宅小区后发现，居民们会严厉地批评那些逃学、制造麻烦、在墙上乱涂乱画的青少年。他们愿意干涉小区内务，维护对小区有利的机构。当地方消防队受到预算削减的威胁时，共同体居民将对此进行干涉。这些都是作者们所称的“集体效力”(collective efficacy) 的例子。但在另一些小区中，居民们则采取一种不干涉的态度。作者们还发现，小区之间的集体效力水平存在着相当大的差异，但不论在富人还是穷人、黑人还是白人的社区里，都可能呈现出或高或低的集体效力水平。值得注意的是，种族异质性在预测低水平的集体效力时，相比于经济劣势、低房屋所有率和其他居住不稳定的指标，显得不那么重要。作者们还发现集体效力水平高的地方，暴力犯罪明显较少。芝加哥住宅小区的例子说明了共同体规范的非正式执行力。

Platteau, Jean-Philippe 和 Seki (1999) 研究了日本富山湾的捕鱼合作组织，阐明了共同体问题的另一方面。面对不确定的捕捉物和所需的较高技术水平，一些渔民选择共享收入、信息和培训。一个 35 年前成立，至今还非常成功的合作组由 7 条捕虾船的全体船员和船长组

成。这些船队共享收入，共担成本，共同修理破损的渔网，共享捕虾地点变换的信息。年长的成员传授他们的技术，受更高教育的年轻成员则教授其他人远距离无线电导航系统和声呐的使用方法。合作组的收入共享和成本共担活动允许船队到更具风险、产量更高的地区捕鱼，共享技术和信息则提高了利润，缩短了船队间生产力在船只、捕鱼、远距离装运和营销上的差距。这与增加利益分享过程的透明度，使得机会主义行为更容易被察觉的过程是同步的。

在俄勒冈州和华盛顿州拥有自己企业的夹板工人吸取了芝加哥住宅小区相互监督方式和渔民风险共担方式两方面的经验。他们选举自己的管理者，要求成员分享企业的所有权并以此作为雇佣的条件，同时又以在企业工作作为拥有企业所有权的条件。在这个产业向美国东南部转移之前，这些合作组已能成功地和传统企业竞争，时间达两代人之久。他们的成功很大程度上要归功于高水平的工作承诺和管理监督成本的节省（当一家企业的性质转变为合作社所有制之后，管理者可以减少3/4）。Ben Craig 和 John Pencavel（1995）通过计量研究指出，合作组的全要素生产率明显高于传统企业。当遇到夹板需求周期性下降的时候，这些合作组并不解雇工人，只是减少他们的工资和工作时间，使得周期性风险不再强加于一小部分人身上，而是由所有共同体成员来分担（参考 Pencavel，2000），还可以在 Hansen（1997），Ghemawat（1995），以及 Knez 和 Simester（1998）那里找到更多的例子。

如这些例子所示，共同体可以解决经典的市场失灵或政府失灵问题：像因为以邻为壑造成地方公共品供给不足，缺少保险和其他分担风险的业务（即使对双方都有利），将穷人排除在信贷市场之外，以及对工作努力过多和无效的监督等等。共同体之所以有时可以弥补政府和市场失灵，是因为共同体中的成员而非外部人拥有其他成员行为、能力和需求的关键信息。共同体成员可以利用这些信息维持共同体规范（例如夹板工人间和渔民间的工作规范、芝加哥共同体的行为规范），同时还可以通过这些信息选择有效的制度安排以避免道德风险和逆向选择问

题。信息分散在共同体内部，一个扬眉的动作，一句话，一个警告，一些闲话或者嘲笑，当它们在人们习惯称“我们”而不是“他们”的邻居或同事之间互相传递的时候，所有这些言行举止都可能含有特殊的意义。

因此，当市场契约和政府指令失效的时候，共同体对治理起了重要的作用，因为法官、政府官员或者其他共同体以外的人无法有效利用那些对加强有利的市场交换和政府指令来说所必需的信息。然而共同体成员可通过彼此间持续的联系，加强信任、互相关心，从而支持群体规范简单有效的执行。这种观点在社会学中早就存在，即使是经济学家中也很早就关注到了这个现象。30 年前，阿罗和德布鲁首次给出亚当·斯密 200 年前关于市场“看不见的手”论断的完整证明。但是福利经济学基本定理所必需的公理要求非常严格，以至于阿罗强调了我们现在称之为社会资本的东西的重要性：

> 缺少信任……人们会失去互利的合作机会……社会行为规范，包括伦理道德准则(可能是)……对弥补市场失灵的社会反应。(Arrow 1971:22)

共同体就是维持这些规范的一种方式（Bowles and Gintis，1999；Bowles and Gintis，1998）。

3　共同体和激励

目前，比较制度分析的任务就是在不考虑计划还是市场的争论的情况下，阐明不同的制度组合能处理哪类问题。契约理论、机制设计、博弈论以及相关领域的发展已经能让经济学家处理相当多的问题。市场之所以吸引人是因为它能利用私人信息。所以当交易者可签订完全契约并能在低成本下执行该契约时，市场通常优于其他的治理结构。

此外，当契约的剩余索取权和控制权紧密相连的时候，市场竞争提供了一种分散的和难以被破坏的机制：惩罚无能者，奖赏高绩效者。

和市场一样，政府也能相对较好地处理某些特殊问题。特别是，对于私人部门之间的交往，政府有权力制定并强制执行博弈规则。所以当参与项目带有强制性时（例如社会保险或建设国防），由政府控制的经济过程最有效。

但是，共同体可以解决一些政府和市场都失灵的问题，尤其是当社会交往的性质或者所交易的商品和服务的性质无法在契约中完全界定，或者其界定成本极为高昂的时候。共同体治理依靠分散的私人信息，根据共同体成员是否遵守社会规范而作出相应的奖励和惩罚，通常这些分散的信息是政府、雇主、银行和其他大型的正式组织难以获得的。一个有效的共同体会监督其成员的行为，使得成员对他们自己的行为负责。与政府和市场相比，共同体能更有效地鼓励和利用激励措施，使得人们按传统规制他们的共同行为，这些激励包括信任、团结、互惠、名誉、自豪、尊重、报复和报答等。

共同体的以下几个方面可解释其作为治理结构特有的能力。第一，在一个共同体中，今天互相接触的成员，今后仍然接触的可能性很高，所以人们有强烈的动机按对社会有利的方式行事，以避免日后他人的报复行为。第二，共同体成员间的频繁交往使得他们彼此间能发现更多的个人信息，了解彼此近期的行为和未来很可能采取的行为，从而降低交易成本、提高收益。共同体成员越容易广泛地获取分散的个人信息，他们就越有动机向有利于集体利益的方向行事。第三，共同体通过其成员直接惩罚“反社会”行为的人克服搭便车问题。当影响他人福利的个体行为不受强制实施的契约约束时，不良动机就会不断滋长，而在工作团队、信用社、合伙经营、地方公共事物和住宅小区中，成员间的互相监督和惩罚是削弱这些不良动机的有效办法。（Whyte，1955；Homans，1961；Ostrom，1990；Tilly，1981；Hossain，1988；Dong and Dow，1993b；Sampson，Raudenbush and Earls，1997）

为了理解共同体的运作机制，经济学家还是从理性个人出发，处理那些一眼看上去不合作肯定是占优策略但最终却形成合作的模型。我们在很多地方已经阐释过为什么我们认为这些解释是不充分的（Bowles and Gintis，2001；Gintis，2000a；Bowles，2002）。与此相比，经济学之外的行为科学家一直通过对利他、情感以及其他非自涉动机的考察来解释共同体。但是，许多研究都缺乏对共同体结构的考察，没有关注共同体结构是否与基于意向性行为的均衡这种传统看法一致。在这个部分里，我们引入一个基于方法论个人主义和面向均衡的经济学模型（或者说，博弈论模型），再加上一种特殊的他涉偏好（other-regarding preference），即在共同体中个体有遵守集体规范的倾向，哪怕自己付出一定的代价。我们引入非自利动机，因为我们坚信要解释共同体是如何通过相互监督执行规范，必须超越传统的个体行为模型。Besley 和 Coate（1995）引入了社会惩罚机制，Kandel 和 Lazear（1992）引入了同辈压力机制，都表现出对传统行为模型的不满。共同体往往有强制执行规范的能力，因为有一部分个体愿意付出一定的代价来惩罚那些卸责的人，虽然他们自身无法得到明显的回报。我们把这种行为称为“强互惠性”。强互惠主义者无条件地与他人合作，并惩罚不合作者，即使惩罚行为不符合传统的理性人假设。鲍尔斯和金迪斯（2000）提供了大量的互惠性证据。也可以参考 Fehr 和 Gächter（2000）以及金迪斯（2000b）。

大量的证据表明，许多社会里面，在不同的社会条件包括彼此完全陌生的情况下，人群中都有相当比例的人是强互惠主义者（Herich，Boyd，Bowles，Camerer，Fehr，Gintis and McElreath，2001）。我们这里回顾一下公共品博弈实验的证据，它的博弈结构与共同体治理的问题非常接近。至于其他相关证据，包括独裁者博弈、最后通牒博弈、共用资源与信任博弈，可以参考 Güth 和 Tietz（1990），Roth（1995）以及 Camerer 和 Thaler（1995）。

公共品博弈由 n 个严格匿名的受试者组成。每一个受试者给予 w

点，实验结束后，点数都可以兑换成现金。每轮博弈中，每个人可以把一部分点数放入一个“共同账户”，保留其余的点数。该轮博弈结束后，每个人除了手里保留的点数，还获得 $q \in (1/n, 1)$ 倍公共账户的点数。于是，为公共账号存钱就变成一种利他行为，因为它增进了团队的收益（$q > 1/n$）而减少了个人的收益（$q < 1$）。

如果受试者是自利的，那么不为公共账号存钱在公共品博弈中就是一种占优策略。但是，在公共品实验中，只有一部分人的行为符合自利的模型。通常情况下，受试者在博弈开始时一般会往公共账户里存入一半的点数。

如果博弈持续几轮，那么为公共账户存钱的点数趋向于减少。在 Fehr 和 Schmidt（1999）的 12 轮公共品博弈实验中，早期几轮大家对公共账户的贡献大约在 40% 到 60% 之间。在最后阶段通常是（第 10 轮），73%的个人（N = 1 042）什么都不捐献，还有很多人的贡献率接近于 0。对此的解释是，受试者发现为公共池捐献还不如不捐，认为自己被那些捐献很少或者不捐的人欺骗了，他们惟一可以用来抵制搭便车的人的办法就是降低自己的贡献率（Andreoni，1995）。实验支持这一解释。当实验允许个人惩罚那些不作贡献的人的时候，他们情愿承担一部分成本来惩罚那些人（Dawes，Orbell and Van de Kragt，1986；Sato，1987；Yamagishi，1988a，b，1992；Ostrom，Walker and Gardner，1992）。

例如，Fehr 和 Gächter（2000）就建立了一个 n = 4 的 10 轮公共品博弈实验，实验中存在有代价的惩罚，并有三种不同的方式分配族群人员。在私人（personal）情境下，4 个受试者组成一个团队不变，进行 10 轮博弈。而在陌生人（stranger）情境下，每一轮受试者都随机搭配。最后，在完全陌生人（perfect stranger）情境下，受试者被随机搭配而且保证以前博弈过的个人永远不会再相遇（在这种情况下，总的博弈轮次必须从 10 轮减少到 6 轮，以避免受试者的重复相遇）。平均来看，每个受试者能够从实验中获得 35 美元的收益。

Fehr 和 Gächter（2000）分别在有惩罚和无惩罚的条件下进行了 10

轮的博弈实验。他们的结果表示在图 1 中。我们可以看到，当有代价的惩罚机制被允许实施时，合作不会最终崩溃。在伙伴博弈的情境下，虽然是严格匿名的，但合作水平持续提高，几乎到达完全合作，哪怕在最后一轮都是如此。但如果不使用惩罚机制，那么这些受试者在进行同样的公共品博弈实验时，合作最终解体了。

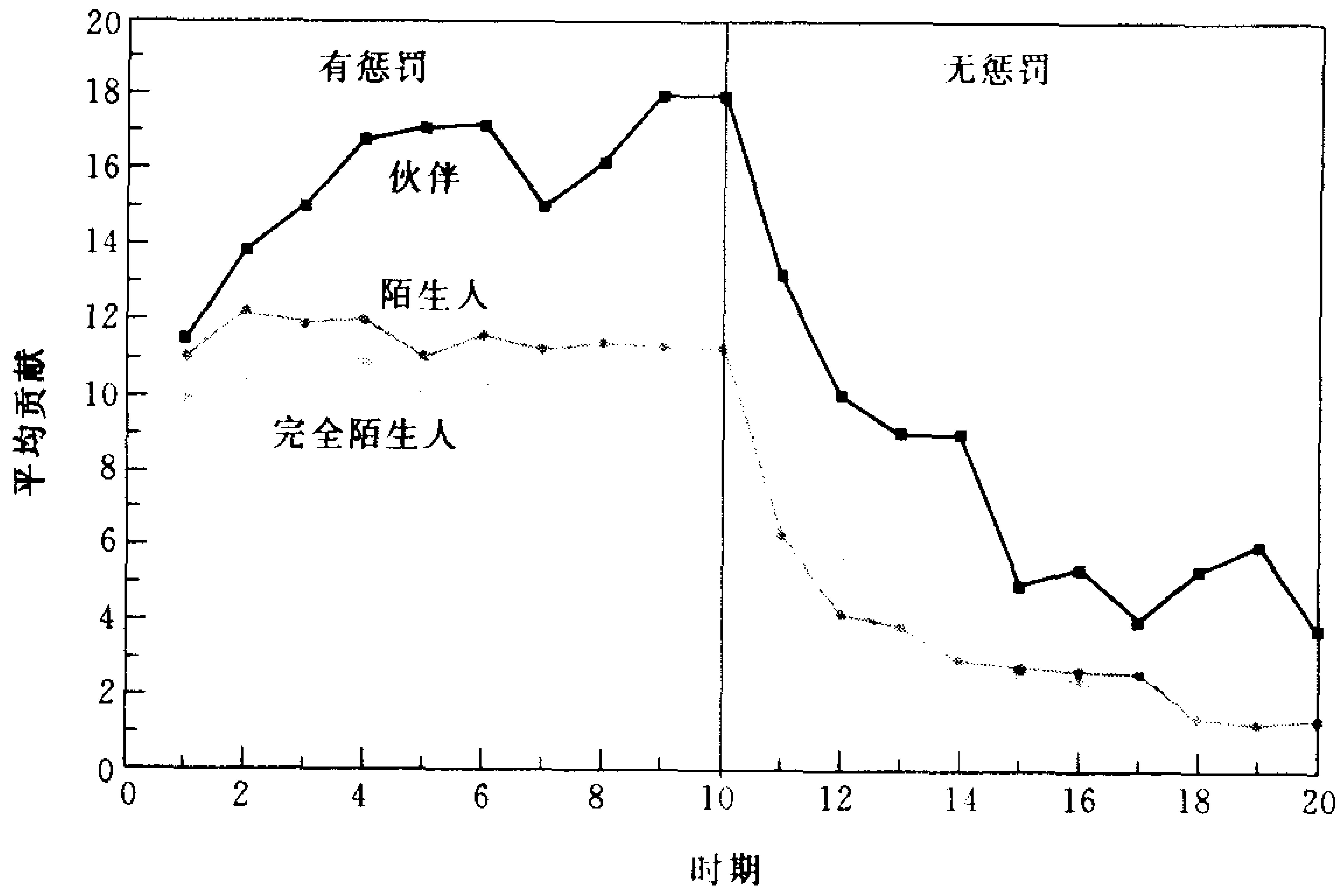

图 1　伙伴，陌生人与完全陌生人环境下，实施惩罚机制的平均贡献领先。(取自 Fehr and Gächter, 2000)

伙伴效应与陌生人关系之间的对比是值得关注的。在后一种情境下，惩罚机制能防止合作的变质。对前者来说，惩罚机制则能持续推动合作比例的上升，直到接近完全合作的程度。这个结果表明，受试者个人有意愿惩罚搭便车者（在陌生人情景中），但在能识别身份的共同体中博弈时，惩罚搭便车者（与伙伴博弈）的意愿更强烈。这样，越是维持强烈的互惠性倾向，共同体就越紧密，存在的时间也越长。

受试者经常自己承担损失来惩罚团队里其他人，这种情况对标准的行为模型提出了重大挑战。在完全陌生人的环境下（或者在其他环境下博弈的最后一轮），占优策略必定是什么都不贡献，也不惩罚别人。事实上，从策略的角度来看，惩罚机制与对公共品作贡献是相等的行

为。惩罚机制是典型利他主义的：做对别人有利的事，自己承担成本。受试者惩罚低贡献者，并且在被询问为什么这样做时表现出强烈的厌恶态度，这些事实都表明这种惩罚是情绪性的反应。更确切地说，这是一种愤怒感的表达。

我们自己的实验（Bowles，Carptenter and Gintis，2001；Bowles and Gintis，2001）表明，在完全陌生人的公共品博弈实验中，惩罚低贡献率者并不仅仅是希望他们能够在下面的博弈中多贡献一些，从而使得大家都能获得更多。同样，被惩罚者在后面的博弈中表现积极也并不是因为他们想避免未来继续受惩罚。我们猜想，可能是惩罚引发了推卸责任者的羞耻感，这使得他们对惩罚作出积极的回应（并不是简单的收益最大化），这是对惩罚有效性的一种解释。公共品博弈实验表明，个人动机包括互惠性偏好可以通过互相监督保持在一个很高的水平上，即使在很大规模的共同体里也是如此（Bowles et al.，2001）。长期以来，经济学忽视这些非标准的动机。不过，如果这些非标准的动机是使共同体运作的部分原因，那么它们也一样隐含着共同体的失灵。

4 共同体失灵

共同体跟市场、政府一样也存在失灵。成员彼此间持久的接触要求共同体的规模要相对小，同时，由于共同体成员偏好与共同体内部的其他成员交往，这经常限制了他们发展从外界更广泛的交易中获得利益的能力；此外，共同体相对同质性的趋势使其不可能获得由不同的知识和其他技术补充所能带来的经济多样性利益。然而这两点缺陷并非无法克服。通过共享信息、设备和技术，日本渔民实现了合作程度低的群体很难达到的规模经济，并从成员才干的差异中获得了实实在在的好处。同样，在被称为“第三个意大利”的地方，当地商业网络和政府一起使得原先无法独立生存的小企业在营销、研究和培训的规模经济中

获益，让它们能与其他的公司巨头竞争。但是与官僚机构和市场相比，共同体有限的规模经常造成不可避免的成本。

共同体失灵的第二个方面比较不明显。当由个体选择而不是由集体决定群体成员时，群体的构成就很有可能会在文化和人口统计上更具同质性，因此剥夺了人们不同的价值观。假设大多数住宅区的人口由仅通过外表和谈吐就能辨识的两类人组成，每个人都强烈地选择加入大群体而不愿意成为少数派。如果个体都想将他们自己归入共同体群体中，那么在全部共同体成员中将产生一种趋势，即所有成员都将最终处于被隔离的状态，其原因在 Thomas Schelling（1978）关于邻里倾卸的文献中作了解释。整合后的共同体能使其中每一个成员的福利获得改善，但是当成员可以自由进入或退出该群体时，这个共同体就无法维系了。Young（1998）和 Bowles（2001）建立模型证明了这个结果。

经济学家使用“市场失灵”和“政府失灵”术语来指出这些治理结构有时会导致资源配置的无效率，我们讨论中所述及的“共同体失灵”也指这种情况。但同政府和市场一样，共同体经常在其他地方，有时甚至是在莫名其妙的情况下失灵。大多数个体在熟悉的人中寻找群体成员，不然他们会感到孤独。但这种归属要求会导致恶待那些不喜欢合群的人，群体的同质性往往会加剧这个问题。当基于种族、地域、民族和性别在道德上不一致而引起内部人和外部人的差异时，共同体治理会比市场和政府更多地导致地方狭隘主义和种族对抗。当内部人有财有权而外部人遭受剥削时，共同体的这个问题将变得更为突出。

这是一个地方性问题。共同体成员有效地工作是因为他们擅长加强规范，并依据规范而确定事情的好坏或对错。开普镇附近地区的白人居民最近反对种族融合，正如人们可想像到的，这是社会资本发挥作用的结果（Jung，1998）。更显著的是 Dov Cohen（1998）对美国暴力和共同体稳定之间关系的区域差别所作的研究。Richard Nisbett（1996）阐述了“荣誉文化”（culture of honor）现象，他发现在美国南部和西部的白人男子中，荣誉文化通常会使人们由公众场合的辱骂和

争吵转变为势不两立的对峙。Cohen 的研究证实在美国北部地区由于争吵而导致的谋杀事件在那些较稳定的住宅区中发生得相对较少。但是这个结论在南部和西部地区正好相反，在荣誉文化占主导的地方，住宅区的稳定与谋杀事件的数量显著正相关。

5 加强共同体治理

很多自由主义哲学传统的追随者，不论是保守的自由放任鼓吹者还是他们的社会民主和自由社会主义批评家，都只把共同体视为远古时期的残留物，它们缺少治理所必需的产权、市场和政府。从这点上说，共同体不是解决市场和政府失灵方案的一部分，而是地方民粹主义或者传统原教旨主义所引出问题的一部分。很多持有这种观点的人一直反对武断地拥护计划或者市场，但他们仍然坚定地把锚抛在好政府这条船上，目前争论的焦点是：最佳的治理结构应落在从政府到市场中的哪一点上。

而那些倡导社会资本的人，或者像我们这样认为共同体治理是政策制定和制度建立的重要方面的人，则对上述观点表示不满，质疑（如 Arrow）政府和市场的组合是否已经很完美以致其他社会规范都已多余。我们相信共同体治理的本质缺陷可以由足够的社会政策来填补。很多学者也指出完善市场或是确保政府干预成功的努力，会破坏虽不完美但并非无价值的以共同体为基础的治理系统，他们认为仅仅将政策范式限定在政府和市场上，可能达不到预期的目标。

我们的看法跟新古典经济学课本上的乌托邦资本主义和乌托邦计划经济的一个分支——福利经济学不同。福利经济学在过去的 50 年中一直天真地认为政府同时拥有获取信息和弥补市场失灵的能力，但事实上并没有理想的共同体治理的蓝图。正如 Ostrom（1990）、James Scott（1998）以及其他领域的研究者所强调的那样，共同体包含上百种不同

的成员规则、事实上的产权形式和决策制定过程，以各种各样的方式来解决问题。但以上这些可能间接地说明了一些经常能在绩效良好的共同体中发现的共同要素，它们也许能够成为旨在加强共同体治理的公共政策的重要部分。

首先，我们的实验证据强烈支持的政策措施是：共同体成员在解决他们共同面对的问题时应当分享成功的利益，分担失败的损失。在日本渔民的例子里，船长和船员都拥有合作组的产出，因而能从成功中直接受益。在芝加哥的住宅小区中，即使控制了大量的人口统计和经济变量，普遍拥有房屋所有权的共同体仍然显示出较高水平的集体效力。最有可能的解释是，房屋所有者从他们与邻居互动关系的改善中完全获益，而不仅仅获得更好的生活质量。这个解释与 Sidney Verba 等（1995）指出的一致，他们认为在控制了大量的人口统计和其他变量后，美国房屋所有者更愿意参与地方而非国家政治。Edward Glaeser 和 Denise Depasquale（1999）也发现德国人房屋所有权的变化预示了公民参与水平的变化。最后，夹板工人获得成功的例子太复杂，用分享合作组收入的剩余索取还无法解释，因为每个人的收益都与其他人的努力结果相关。总体看来，这些例子说明，为实现成功，共同体成员必须拥有那些资产，他们或者使用其工作，或者以共同体工作影响其价值。

其次，正如我们已经在带有惩罚的公共品实验中看到的，如果监督机制和惩罚不合作者的机制能够内化进入互动的结构中去，那些会颠覆共同体的破坏合作机制的行为就能被制止。那些旨在提高共同体内个体互动行为透明度的政策，再加上鼓励对卸责者采取各种形式惩罚的机制，就能解决合作的问题，即使多数个体都是自利的。狩猎—采集的原始部落往往保持一种风俗：大家一起分食所得的食物。这种机制的目的就是使大家都能很容易地发现谁违反了分配规则。日本富山海湾渔民在统一时间上岸卸货，以保证实施公平的分配机制。

在这些描述适度规模团队内合作行为的模型里，一个非常重要的特征是，通过惩罚卸责者而维持合作的多种均衡存在。如果合作是普

遍的，那么具有规范意识的人惩罚卸责者的成本很小，规范就可以维持下去。当合作非常罕见时，那些惩罚卸责者的人本身要付出极高的代价，从而使得他们在演化过程中被淘汰（Boyd，Gintis，Bowles and Richerson，2001）。这就意味着，在具有文化规范的异质人群中，是否去惩罚那些违背社会规范或者自私自利的人从而表现出高水平或者低水平的合作率，不仅与人群中不同类型人的分布有关，更与团队近期的历史相关。

早在1754年，休谟就说过这样的话（Hume，1898［1754］）："那就是说，为了设计任何治理体系……每个人都必须被假定为流氓。在他一切行为中，除了私利，再没有其他目的。"但休谟是诉诸于审慎，而不是现实。他下一句话就说："奇怪的是，这句格言在政治中应该是正确的，可事实上是错的。"如同休谟意识到的那样，个人并不仅仅是自私的，常常还有关注荣誉等情感。于是，审慎和现实的关系如同一句格言所说的，政策和宪法制订者应该知道人是异质的。个体能够被朝各个方向塑造和改变。好的政策和宪法不仅要驾驭个人自私动机，使之走向社会价值观目标，还能够唤起、培育人们的公共精神。

第三个必要条件是，上述例子以及其他很多类似的例子都表明：运作良好的共同体需要一个法律和政府环境支持其发挥功能。芝加哥居民小区里如果没有随时待命的警察，其减少犯罪率的成果就很难实现。日本的超过1 000名渔民的合作组，是在国家和地方环境以及其他规制下实现的，他们可以自由地制订地方规范与国家规范互补，但不会越界。学者们比较台湾地区和印度南部农民的灌溉组织，发现前者的成功应归功于政府的有效干预、提供有利的法制环境、处理那些共同体的非正式规定不够有效的事物（Lam，1996；Wade，1988）。类似的"共同体—政府"协同作用在Tendler（1997）关于获取卫生保健的研究和Ostrom（1996）关于城市基础设施建设的研究中有所论述，二者都是关于巴西的研究。政府干预有时会破坏共同体治理能力，但这并不意味着我们应该支持极端自由主义。

所以，面对面的局部互动并不是有效治理结构的替代物，而只是一种补充。社会资本这一概念的普及很可能是因为人们忽视了这一点。最近的一项盖洛普民意测验中，实验者询问了大量美国人，“你认为以下哪一个群体对帮助穷人负有最大的责任：教会、私人慈善机构、政府、穷人的家庭成员和亲戚、穷人自己还是其他人？”他们同时也询问了收入和财富不平等是否“可接受”或者是否“是一个需要解决的问题”。测验结果表明，样本均匀地分为两种情况，一种认为政府应当对此作出反应，而另一种则认为应由非政府组织来解决；在那些漠不关心收入和福利水平差距的被访者中，支持以私人方式来解决问题的人3倍于选择政府来解决问题的人。[2] 在该测试中，那些选择社会资本相关选项的人可能更倾向于认为社会资本可以减少政府干预，而并非希望靠它来减少不平等。

因此，与共同体治理能力互补的法律和政府环境，以及能保证成员从共同体成功中获利的产权分配制度都是加强共同体治理的关键因素。发展出使得政府、市场和共同体三者互补的制度框架是一项极具挑战性的任务。例如，当模糊的产权和非正式契约的执行对互惠交易必不可少时，更精确界定的产权可能会降低共同体成员间接触的多面性和重复性（Bowles and Gintis，1998）。同样，大量证据表明，利用惩罚和认知机制可以调动自利的动机，以达到较高的工作努力、遵守规范和保护环境等水平，但这也有可能会破坏互惠行为以及其他社会动机（Fehr and Gächter，2000；Bewley，1995；Gneezy and Rustichini，2000；Cardenas，Stranlund and Willis，2000；以及 Bowles 1998 所引用过的其他资料）。

共同体有效治理的第四个要素是：倡导平等对待的自由道德观念和加强反歧视政策。成员间没有“我们”敌视“他们”的行为，共同体能够达到有效治理，这并非不现实。很多运作良好的共同体都能说明这点，这些共同体在治理过程中并没有表现出我们前所述及的丑陋的地方观念和潜在分裂趋势。

我们还能设想出其他加强共同体治理的方法，但是有些措施增加了

在共同体有效治理和前面提到的地方观念之间权衡比较的困难。例如，Alesina 和 La Ferrara（1999）考察了美国一些地区，发现当控制其他众多可能的影响因素后，在收入分配越公平的地方，成员参与教会、地区服务和政治组织以及其他共同体组织的水平也越高。他们的调查显示，增加收入平等的政策将加强共同体治理。此外，他们在共同体中随机选取两名成员，测量这两个人属于不同种族的概率，发现种族差异大的地区成员的参与水平显著较低。因此，人们也许认为有利于共同体的公共政策将不会增加群体中的种族同质性。

但是简单地反对同质化治理政策的理由也是不充分的。Alesina 和 La Ferrara 的研究结果以及其他学者类似的研究也表明，成功的共同体很有可能是同质性较高的，那么对共同体治理的高度信任在缺少足够的制衡政策时会促使高水平的地方同质化，因为族群的成功以及他们可能存在的时间将随着同质程度的变化而变化。因此，工人集体所有制企业组成的竞争性经济实体可能比传统经济体表现出更多的同质性。族群内部的同质性和族群之间的竞争性这两者之间的关系，有效地促进了好的治理形式的产生，从而减轻了“我们对他们”这种敌对情感。但是这两者之间的矛盾不可能完全消失。

6 经济演化和共同体治理的未来

人们普遍认为商业时代和民主政治的到来将使共同体黯然失色。持各种信仰的作者都认为市场、政府或者简单地说“现代化”将毁灭一种价值观念，这种观念在整个历史过程中维持着一种基于亲密情感的治理形式。正如浪漫派保守主义者埃德蒙·伯克（1955［1790］）所说：

> ……骑士时代已经一去不复返了。那些诡辩家、经济学家和精于计算的人们胜利了……没留下任何对国家集体的感情……

以在我们中间产生友爱、尊敬、钦佩或是眷恋。

自由主义者托克维尔通过回顾1830年间美国的民主文化，对伯克的忧虑作出回应：

> 每一个人……对于其他人而言都是陌生人……他的孩子和密友是他全部的人生；至于身边的其他人，他接近他们但却看不到……接触他们但又感觉不到；他只为他自己而存在……

社会主义者马克思和恩格斯（1972［1848］）提出：

> 资产阶级……把一切封建的、宗法的和田园诗般的关系都破坏了。它无情地斩断了把人们束缚于天然尊长的形形色色的封建羁绊，它使人和人之间除了赤裸裸的利害关系，除了冷酷无情的"现金交易"，就再也没有任何别的联系了。……在无数争取自由的地方，资本主义建立了单一的、不合理的自由——自由贸易。

很多预言共同体消失的学者将他们的论据建立在这样的想法上，即共同体的存在应归功于一套独特的前现代（pre-modern）价值观，这种观念一定会被市场中的经济竞争和民主政府中的政治竞争所磨灭。现代作者也强调地方观念需要文化认同，这些认同与现代社会制度相对立。Talcott Parsons 的社会学体系提出"特定的"（particularistic）价值观与原始文明相一致，而"普适的"（universalistic）价值观则与近代文明相一致。

Fred Hirsch 同样认为前资本主义的道德准则正逐渐减弱：

> 这种遗产随着时间和资本主义价值观的腐蚀不断减少。个

体行为越来越向对自己有利的方向发展，建立在共同态度和目标上的习惯与本能已经丢失了。Hirsch(1976):117—118。

我们并不怀疑市场和民主政府塑造了文化环境，提升某些价值观同时打击另一些。我们确实可以预见到伯克、马克思和托克维尔在很久以前就指出的价值观念的衰落。但是共同体发展、衰落和转换的基础并不是早期残留的价值观，而在于共同体的能力，像市场和政府一样能成功解决这个时代的社会协调问题的能力。

共同体治理并不是过时的治理方式，它在未来社会可能会变得更重要。原因在于，当个体间互动行为过于复杂或交易信息无法核实，使得完全契约或外部命令难以约束个体之间的行为时，那些需要由共同体解决，同时政府和市场都难以解决的问题会越来越多。这类个体间的互动行为在现代经济社会中日益增多，诸如信息密集的团队生产逐渐取代装配线生产以及其他用契约和指令更容易处理的生产技术，还有像难以测量投入和产出的服务等等。在日益以质量而不是以数量为目标的经济体中，共同体通过互相监督、共担风险和共享收益机制逐渐表现出优秀的治理能力。

但是共同体解决问题的能力可能受制于成员间等级差异以及经济上的不平等。许多观察家相信，日本企业中管理者和工人之间有限的不平等是管理人员和生产工人之间信息分享的关键因素（Aoki，1988）。Dayton Johnson 和 Bardhan （2002）发现在印度和墨西哥等地的灌溉组织中，如果农民之间存在有限的身份和等级不平等，那么他们就较有可能合作以有效利用水资源。这些研究结果反映出的行为规律与以下实验结果相同，即当隐含在收益矩阵中的利益冲突程度增强时，在两人不重复的囚徒困境博弈中的合作水平将急剧下降（Axelrod，1970；Rapoport and Chammah，1965）。

如果我们以下的论述是正确的，即当任务只是定性的且难以明确地在契约中表达，同时成员间的利益冲突又比较有限，共同体就能够良好

地运作，那么极端不平等的社会很有可能将来会处于竞争劣势，因为他们的特权结构和物质报酬结构限制了共同体治理在现代经济系统下促进定性互动行为的能力。

注释：

[1] Ouchi (1980), Hayami (1989), Ostrom (1997), Aoki 与 Hayami (2000) 都提出过类似的看法。

[2] 来自 Christina Fong (1999) 对盖洛普民意测验社会审计调查 (Gallop Poll Social Audit Survey) 中的数据分析结果，这项名为“富人和穷人：对公平和机会的理解” (Haves and Have-Nots: Perceptions of Fairness and Opportunity) 的调查于 1998 年 4 月 23 日到 5 月 31 日在全美展开，调查随机抽取了 5 001 名成人作为样本。

参考文献：

Alesina, Alberto and Eliana La Ferrara, “Participation in Heterogeneous Communities,” 1999. NBER Working Paper 7155.

Andreoni, James, “Cooperation in Public Goods Experiments: Kindness or Confusion,” *American Economic Review* 85, 4 (1995): 891—904.

Aoki, Masahiko, *Information, Information, Incentives, and Bargaining in the Japanese Economy* (Cambridge: Cambridge University Press, 1988).

— and Yujiro Hayami, “Introduction,” in Masahiko Aoki and Yujiro Hayami (eds.) *Communities and Markets in Economic Development* (Oxford: Oxford University Press, 2000).

Arrow, Kenneth J., “Political and Economic Evaluation of Social Effects and Externalities,” in M. D. Intriligator (ed.) *Frontiers of Quantitative Economics* (Amsterdam: North Holland, 1971) pp.3—23.

Axelrod, Robert, *Conflict of Interest: A Theory of Divergent Goals with Applications to Politics* (Chicago: Markham, 1970).

Baland, Jean Marie, Pranab Bardhan and Samuel Bowles, *Inequality, Cooperation and Environmental Sustainability* (New York: Russell Sage, 2002).

Bardhan, Pranab, Samuel Bowles, and Herbert Gintis, “Wealth Inequality, Credit Constraints, and Economic Performance,” in Anthony Atkinson and Francois Bourguignon (eds.) *Handbook of Income Distribution* (Dortrecht: North- Holland, 2000).

Besley, Timothy and Stephen Coate, “Group Lending, Repayment Incentives and Social Collateral,” *Journal of Development Economics* 46 (1995): 1—18.

Bewley, Truman F., “A Depressed Labor Market as Explained by Participants,” *American Economic Review* 85, 2 (1995): 250—254.

Bowles, Samuel, “Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions,” *Journal of Economic Literature* 36 (March 1998): 75—111.

— *Economic Behavior and Institutions: An Evolutionary Approach to Microeconomics*, 2002. Princeton University Press.

— and Herbert Gintis, “The Moral Economy of Community: Structured Populations and the Evolution of Prosocial Norms,” *Evolution & Human Behavior* 19, 1 (January 1998): 3—25.

— and —, *Recasting Egalitarianism: New Rules for Markets, States, and Communities* (London: Verso, 1999). Erik Olin Wright (ed.).

— and —, “Walrasian Economics in Retrospect,” *Quarterly Journal of Economics*

(November 2000).

— and —, "The Economics of Shame and Punishment," 2001. Santa Fe Institute Working Paper.

—, Jeffrey Carptenter and Herbert Gintis, "Mutual Monitoring in Teams: The Importance of Shame and Punishment," 2001. University of Massachusetts.

Boyd, Robert, Herbert Gintis, Samuel Bowles and Peter J. Richerson, "Altruistic Punishment in Large Groups Evolves by Interdemic Group Selection," 2001. Working Paper.

Burke, Edmund, *Reflections on the Civil War in France* (New York: Bobbs-Merrill, 1955 [1790]).

Camerer, Colin and Richard Thaler, "Ultimatums, Dictators, and Manners," *Journal of Economic Perspectives* 9, 2 (1995): 209—219.

Cardenas, Juan Camilo, John K. Stranlund and Cleve E. Willis, "Local Environmental Control and Institutional Crowding-out," *World Development* 28, 10 (July 2000).

Cohen, Dov, "Culture, Social Organization, and Patterns of Violence," *Journal of Personality and Social Psychology* 75, 2 (1998): 408—419.

Craig, Ben and John Pencavel, "Participation and Productivity: A Comparison of Worker Cooperatives and Conventional Firms in the Plywood Industry," *Brookings Papers: Microeconomics* (1995): 121—160.

Dawes, Robyn M., John M. Orbell and J. C. Van de Kragt, "Organizing Groups for Collective Action," *American Political Science Review* 80 (December 1986): 1171—1185.

Dayton-Johnson, J. and Pranab Bardhan, "Inequality and the Governance of Water Resources in Mexico and South India," in Jean Marie Baland, Pranab Bardhan and Samuel Bowles (eds.) *Inequality, Cooperation and Environmental Sustainability* (New York: Russell Sage, 2002).

de Tocqueville, Alexis, *Democracy in America*, *Volume II* (New York NY: Vintage, 1958).

Dong, Xiao-yuan and Gregory Dow, "Monitoring Costs in Chinese Agricultural Teams," *Journal of Political Economy* 101, 3 (1993): 539—553.

Durlauf, Steven, "On the Empirics of Social Capital," *Economic Journal* (forthcoming).

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114 (August 1999): 817—868.

— and Simon Gächter, "Cooperation and Punishment," *American Economic Review* 90, 4 (September 2000).

— and —, "Altruistic Punishment in Humans," *Nature* (forthcoming).

Fukuyama, Francis, *The Social Virtues and the Creation of Prosperity* (New York: Free Press, 1995).

Ghemawat, Pankaj, "Competitive Advantage and Internal Organization: Nucor Revisited," *Journal of Economic and Management Strategy* 3, 4 (Winter 1995): 685—717.

Gintis, Herbert, *Game Theory Evolving* (Princeton, NJ: Princeton University Press, 2000).

—, "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology* 206 (2000): 169—179.

Glaeser, Edward, David Laibson and Bruce Sacerdote, "The Economic Approach to Social Capital," *Economic Journal* (2002).

—, —, Jose A. Scheinkman and Christine L. Soutter, "Measuring Trust," *Quarterly Journal of Economics* 65 (2000): 622—846.

Glaeser, Edward L. and Denise DiPasquale, "Incentives and Social Capital: Are Homeowners Better Citizens?," *Journal of Urban Economics* 45, 2 (1999):

354—384.

Gneezy, Uri and Aldo Rustichini, "A Fine is a Price," *Journal of Legal Studies* 29 (2000): 1—17.

Güth, Werner and Reinhard Tietz, "Ultimatum Bargaining Behavior: A Survey and Comparison of Experimental Results," *Journal of Economic Psychology* 11 (1990): 417—449.

Hansen, Daniel G., "Individual Responses to a Group Incentive," *Industrial and Labor Relations Review* 51, 1 (October 1997): 37—49.

Hayami, Yujiro, "Community, Market and State," in A. Maunder and A. Valdes (eds.) *Agriculture and Governments in an Independent World* (Amherst, MA: Gower, 1989) pp.3—14.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath, "Cooperation, Reciprocity and Punishment in Fifteen Small-scale Societies," *American Economic Review* 91 (May 2001): 73—78.

Hirsch, Fred, *Social Limits to Growth* (Cambridge, MA: Harvard University Press, 1976).

Homans, George, *Social Behavior: Its Elementary Forms* (New York: Harcourt Brace, 1961).

Hossain, M., "Credit for Alleviation of Rural Poverty: the Grameen Bank in Bangladesh," 1988. International Food Policy Research Institute Report 65.

Hume, David, *Essays: Moral, Political and Literary* (London: Longmans, Green, 1898 [1754]).

Jung, Courtney, "Community is the Foundation of Democracy: But what if your Community Looks Like This?" 1998. Yale University.

Kandel, Eugene and Edward P. Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy* 100, 4 (August 1992): 801—817.

Knez, Marc and Duncan Simester, "Firm-wide Incentives and Mutual Monitoring," September 1998. Graduate School of Business, University of Chicago.

Lam, Wai Fung, "Institutional Design of Public Agencies and Coproduction: A study of Irrigation Associations in Taiwan," *World Development* 24, 6 (1996): 1039—1054.

Marx, Karl and Friedrich Engels, "The Communist Manifesto," in Robert Tucker (ed.) *The Marx-Engels Reader, 2nd Edition* (New York: W. W. Norton & Company, 1972 [1848]).

Nisbett, Richard E. and Dov Cohen, *Culture of Honor: The Psychology of Violence in the South* (Boulder: Westview Press, 1996).

Ostrom, Elinor, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge, UK: Cambridge University Press, 1990).

—, "Crossing the Great Divide: Coproduction, Synergy, and Development," *World Development* 24, 6 (1996): 1073—1087.

—, "The Comparative Study of Public Economies," 1997. Workshop in Political Theory and Policy Analysis: Center for the Study of Institutes, Population and Environmental Change, Indiana University.

—, James Walker and Roy Gardner, "Covenants with and without a Sword: Self-Governance Is Possible," *American Political Science Review* 86, 2 (June 1992): 404—417.

Ouchi, William, "Markets Bureaucracies and Clans," *Administrative Sciences Quarterly* 25 (March 1980): 129—141.

Pencavel, John, "Worker Participation: Lessons from the Worker Co-ops of the Pacific North-West," June 2001. Stanford University, Department of Economics.

Platteau, Jean-Philippe and Erika Seki, "Community Arrangements to Overcome Market Failure: Pooling Groups in Japanese Fisheries," in M. Hayami and Y. Hayami (eds.) *Communities and Markets in Economic Development* (Oxford: Oxford Univer-

sity Press, 2001) pp. 344—402.

Rapoport, Anatol and Albert Chammah, *Prisoner's Dilemma* (Ann Arbor, MI: University of Michigan Press, 1965) .

Roth, Alvin, "Bargaining Experiments," in John Kagel and Alvin Roth (eds.) *The Handbook of Experimental Economics* (Princeton, NJ: Princeton University Press, 1995) .

Sampson, Robert J., Stephen W. Raudenbush and Felton Earls, "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy," *Science* 277 (August 15, 1997) : 918—924.

Sato, Kaori, "Distribution and the Cost of Maintaining Common Property Resources," *Journal of Experimental Social Psychology* 23 (January 1987) : 19—31.

Schelling, Thomas C., *Micromotives and Macrobehavior* (NewYork: W. W. Norton & Co, 1978) .

Scott, James, *Seeing Like A State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven: Yale University Press, 1998) .

Tendler, Judith, *Good Government in the Tropics* (Baltimore: Johns Hopkins, 1997) .

Tilly, Charles, "Charivaris, Repertoires and Urban Politics," in John M. Merriman (ed.) *French Cities in the Nineteenth Century* (New York: Holmes and Meier, 1981) pp. 73—91.

Verba, Sidney, Kay Lehman Schlozman and Henry Brady, *Voice and Equality: Civic Voluntarism in American Politics* (Cambridge, MA: Harvard University Press, 1995) .

Wade, Robert, "Why Some Indian Villages Cooperate," *Economic and Political Weekly* 33 (April 16 1988) : 773—776.

Whyte, William F., *Money and Motivation* (New York: Harper & Row, 1955) .

Yamagishi, Toshio, "The Provision of a Sanctioning System in the United States and Japan," *Social Psychology Quarterly* 51, 3 (1988) : 265—271.

—, "Seriousness of Social Dilemmas and the Provision of a Sanctioning System," *Social Psychology Quarterly* 51, 1 (1988) : 32—42.

—, "Group Size and the Provision of a Sanctioning System in a Social Dilemma," in W. B. G. Liebrand, David M. Messick, and H. A. M. Wilke (eds.) *Social Dilemmas: Theoretical Issues and Research Findings* (Oxford: Pergamon Press, 1992) pp. 267—287.

Young, H. Peyton, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (Princeton, NJ: Princeton University Press, 1998) .

共同体的道德经济：结构化的人群和趋社会规范的演化*

萨缪·鲍尔斯　赫伯特·金迪斯

1 引　　言

如果我们是正确的，那么共同体的兴起、消亡和变革，并不是由早些时期所遗留下来的价值观，而是由共同体的能力所造成的。正如市场和国家一样，共同体为人们在社会生活中所遇到的问题提供了成功的解决方法。我们所说的“共同体”，是指一种社会互动结构，具有很高的进入和退出成本，并且成员之间相互认识。正如生物学中的“族群”一样，共同体内部成员之间的交往比和外部成员[1]的交往更加频繁和广泛。共同体像族群，但是它不像公司和家庭，因为它缺少一个以其成员[2]为基础的中央集权式的决策机构。共同体的例子有：邻居、老男生网络（old boy networks）、种族协会以及很多商业、贸易和工匠组织。

我们认为“规范”的含义是一种文化特征，能够操纵影响到其他人

* 原文题目为 The Moral Economy of Communities：Structured Populations and the Evolution of Pro-social Norms，发表于 *Evolution and Human Behavior* 19，1（1998）：3—25，林山水译。

的福利的人类行为，同时这些行为不能被无成本的强制性契约所限制。规范的其他用法也曾经被提出来。但是，我们要强调的问题是，既然社会互动会培养出趋社会的受规范支配的行为，那么，社会互动又是如何被构造出来的？趋社会规范就是那些不断提升一个人群的平均福利水平的规范。趋社会规范的例子有：诚实；在“鹰—鸽”互动中使用“鸽式”行为；在囚徒困境下倾向于协调（不管是单方的还是有条件的）；对有反社会行为的人倾向于报复。

趋社会规范的重要性体现在交往过程中所带来的结果。个体的协调行动所产生的结果，优于个体的不协调行动（学术上称为“不合作”）所产生的结果。这样的例子有囚徒困境、鹰鸽博弈以及其他一些多重均衡下的互动（这些均衡中的一部分明显优于另一些）。这些互动一般称为协调问题，而互动过程中所产生的不良结果则称为“协调失败”。协调失败的一般原因是，激发个体行为的收益和成本没有把行为对其他人产生的相关收益或有成本的后果考虑进去。

共同体通过支持与趋社会规范相一致的行为，如诚实、互惠、朝着共同目标合作等方式克服搭便车问题，同时也惩罚了“反社会”行为。这些规范通常被认为是传统文化的历史遗产，通过有意向的教化和人们的普遍认可而得到支持。但是，这种以共同体为基础的规范是不具有强制性的。首先，族群对很多重要的规范具有不同的看法，并且，在说明族群成员的规范性定向（normative orientations）分布时，普遍规范的偏离理论并不能提供足够的解释力（Gintis，1975）。第二，有意向的教导规范所带来的潜在力量被很多失败的社会工程学实验掩饰了。在前苏联和其他地方，试图建立“新社会主义人”就是一个明显的例子。第三，价值取向看上去受制于快速的转变，可以观察到的例子有，本土文化的快速解体以及20世纪现代女权运动的剧烈上演，这表明历史尽管重要，但在当代过程中特殊规范也可以维持。

有一种看法认为以共同体为基础的价值观是过往社会反复灌输下来的遗产，作为对这种思想的替代，我们发展了一种新的观点。我们认

为，社会互动的当代结构可以用来刻画共同体的特征，这种结构不是由传统的内在力量或者有意识的教导所形成的，而是由我们前面所提到的趋社会规范发展而来。“社会互动的结构”指的是这样一种规则，它可以支配人们的需求，决定人们的日常行为，以及这些行为所带来的结果。在本文的后面，我们将会利用一组与群体成员相匹配的规则和通过描述他们配对交往的博弈结构来使社会互动结构形式化。

本文的论证可以概括如下。第一，共同体会影响规范的演化，因为共同体构建的社会互动影响了受规范支配行为的收益和成本，并且规范的获取和保持受到了相关收益的影响。第二，共同体支持了具有高频率的趋社会规范的均衡。我们认为，共同体之所以存在，是因为它们减少了市场、国家和其他竞争性的制度所难以处理的协调失败问题。[3]

因此，我们就把论证重点集中在规范和人们在做事获取规范的方式上面。因为我们所讨论的特征很大一部分是道德准则或行为规则，而其他特征，如一个人的口音，是不能被用来当作研究对象的。我们需要一种比标准经济学更加一般化的方法。在标准经济学那里，行为或者控制行为的规则，是实现目标函数最大化的工具。在本文中，我们引进了演化的观点，这种观点有助于我们理解各种各样的社会互动行为。我们所研究的是差异复制（differential replication）：具有耐久性的行为（包括规范）被模仿、保留、传播和复制，而其他特征则没有这些功能。[4]

异质复制可能来源于个体想要取得和保留那些成功人士所具有的特征。异质复制也可能产生于工具味较弱的原因：那些拥有“成功特征”的人可能是拥有特权的文化典范，如父母或者教师。在权力的应用过程中，异质复制可能发生作用，如在国家、阶级或者其他集体斗争中，失败的一方总是引进成功一方的文化、制度和其他因素。(Kelly，1985；Weber，1976)

2 共同体治理

正如我们所定义的共同体概念，共同体可以建立起社会的互动结

构，这主要是通过鼓励以下内容来实现的：（a）行为者之间频繁的交往；（b）在共同体成员之间，信息成本是比较低廉的；（c）共同体成员倾向于内部互动；（d）共同体之间的移民受到限制。这些结构特征对共同体促进趋社会行为的发展有着重要作用。

共同体的互动结构与市场和国家相反，至少是与它们的理想化形式相反。市场互动的特征是短期的接触，并且互动者是匿名的，而理想化的国家官僚机构的特征是长期的匿名关系。在图 1 中我们进行了相应的对比。作为治理结构，国家和市场都有自己独特的地位和缺点，但在这里，我们关心的是共同体。[5]

	短期的	长期的
匿名的（anonymous）	市场	国家
私已的（personal）	——	共同体

图 1　不同制度下的互动结构

具体来考虑，一个特定的共同体面临着囚徒困境式的协调问题。假设一个共同体是由大量配对交往的人们组成，他们可能的行为和收益如图 2 所示。在常见的收益下：[6]

$$a > b > c > d \text{ 且 } a + d < 2b \quad (1)$$

每个人所选择的行为并不受强制性的契约限制。在这种互动过程中，背叛是最普遍的占优策略均衡。

	协调	背叛
协调	b，b	d，a
背叛	a，d	c，c

图 2　囚徒困境：收益(行，列)

注意：$a > b > c > d$，$a + d < 2b$

如果行为者能够以某种契约的方式进行协调，那么他们当然会选择合作。但是，我们假设行为者是在没有契约限制的情况下进行博弈，也就是说它们的互动是非合作的。虽然如此，共同体的结构为什么会导致普遍的合作呢？我们有三个选择来解决共同体的合作问题。每个选择都是以众所周知的博弈模型为基础的。

第一，共同体成员之间的频繁互动降低了信息收集成本，并且，提高了发现与别人交往可以降低信息收集成本的特点的成员的收益。信息的获取和传播越快，共同体的成员就越有动力，他们会以有利于他们邻居的方式进行活动。因此，在一个重复的互动过程中，他们就会有动机以合作行为建立“声誉”（Shapiro，1983；Gintis，1989；Kreps，1990）。这就是共同体的声誉效应。

第二，在一个共同体当中，如果成员之间今天进行交往，那么他们在未来进行交往的概率就会比较高，因此他们就会以比较善意的方式进行交往，以避免对方在今后出现背叛行为（Axelrod and Hamilton，1981；Axelrod，1984；Taylor，1987；Fudenberg and Maskin，1986）。成员之间的互动越全面，那么惩罚机会主义者的机会就越多。我们把这种效应称为报复效应。

第三，趋社会和反社会的行为者一般会包含有对他人的赠与收益和施加成本，在这种情况下，收益和成本并不服从成本效率规律。在一个人口众多的共同体当中，其成员之间互动的可能性比较大，这也就增加了互动的频繁程度。趋社会行为者更有可能得到奖励，他们之间也更加容易进行交往。对于反社会行为者来讲，则情况刚好相反（Grafen，1979；Axelrod and Hamilton，1981；Bowles，1996）。这就是所谓的分割效应。

上面提到的声誉效应、报复效应和分割效应使得共同体会支持趋社会特征所带来的均衡。由于族群之间的流动性有限，进入和退出的成本比较高，因此这些效应得到了强化。地方观念效应的实现并不是通

过趋社会行为者直接操作，而是通过对声誉、报复和分割效应的强化来实现的（Bowles and Gintis，1997）。注意地方观念效应和族群选择机制不同，它是由族群之间的各种不同特征决定的。在族群内，声誉、报复和分割效应得到了纳什均衡的支持，因此，当族群之间的竞争[7]消失时，这三种效应也是存在的。

图 3 总结了共同体结构和协调问题之间的因果关系。在本文接下来的部分，我们将会来探讨这四种效应的机理，但我们首先必须明确的一点是，与特征相联系的收益是如何影响异质复制的，因为这种关系是理解制度效应对文化演化所产生影响的关键。

	有助于解决协调问题的效应	该效应所必需的共同体特征	模型变量
声誉	对趋社会行为者的强化的声誉价值	其他行为者的低信息成本	δ
报复	对反社会行为者的惩罚	交往的长期性或者频繁性	ρ
分割	反社会行为者的不利配对	行为者的非随机配对	σ
狭隘主义制度	有利于趋社会特征的强化压力	族群间的有限迁徙	μ

图 3　共同体如何解决协调问题

3　经济制度和文化演化

支配社会生活的经济制度和其他规则影响着人群的社会互动结构，互动结构反过来又影响与特定行为相关的收益，这个特定行为受规范和其他文化特征的支配。由于这些收益影响着对文化特征的差异化采纳、保持和放弃，因此制度影响着群体文化特征的均衡分布。也就是说，制度混合的变化影响着文化的演化，这种影响是通过改变社会互动

结构并因此改变文化传播过程来实现的。

我们用一系列的文化特征来定义文化这个词。文化特征是一种信念、价值观或者其他对个体行为方式产生长期影响的后天特征。[8] 乐于助人、喜欢建立起大的家庭、不吃早餐等都是文化的特性，当然，互惠的社会邀请和竞标中高价者有权购买也是文化特征。文化演化是文化特征在人群中的分布随着时间变化的过程。文化均衡是指文化特征的分布不受内生因素变化的影响。文化环境是任何一种影响现有文化特征被其他人接受和保留（不管是自愿、自觉还是不自愿、不自觉）以及新文化特征被引进的社会情境。

共同体就像市场和国家一样，是一种使文化特征得到发展或者改变的环境。不同的文化环境可以通过喜欢复制和发展不同的文化特征来加以区别。[9] 我们假设文化演化就是在群体成员认可的特征异质复制的影响下发生的。[10] 个体文化特征的模仿被认为是社会成功的，这就像生物学上的适存生物体繁殖成功一样。如果成功的文化特征没有道德的压力，并且只是由于其预期结果而得到信奉，那么模仿就会非常迅速。甚至，当一个受到深度信奉的价值观与成功人士的行为发生冲突的时候，这种价值观也可能受到人们的抛弃。这可能发生于族群选择的过程中，这是因为，拥有成功价值观的族群可能取代那些没有成功价值观的族群（Soltis et al.，1995）。另外，个体本身也有可能抛弃某些不合适的价值观（Festinger，1957），或者，新生一代可能拒绝接受老一辈的价值观（Fromm and Maccoby，1970）。除此之外，在一个社会环境（如工作车间）中有用的价值观也可能不自觉地“传递”给其他人，从而威胁或取代传统的价值观（Kohn，1969；Bowles and Gintis，1976）。最后，某些成功的个体，如政府领导人、公众人物、教师等等，他们作为文化的典范有特殊的渠道接近大众，因此，他们对公众的影响力就比较大，他们的文化特征因为其在社会结构中的位置而更容易被模仿（LeVine，1966）。

我们用异质复制的学习规则取代了自觉的最优化。我们并不特别

指出特征为什么被复制。这个课题仍然是开放的。我们只是要指出，成功的特征更容易被复制。

文化的传播是以成功特征被复制为基础的，我们用下面的模型来进行解释。假设有两个相互排斥的特征（x 和 y），它们代表人群中的每个成员（没有 x 时，y 存在）。用 p_x 表示人群中拥有 x 特征的成员，γ_x 表示 p_x 随时间增长的比率。[11] 文化传播过程的结果是这样的：在一个大量人口的群体中，给定特征的分布，由行为者来执行由自己所分配到的特征决定策略，在这种情况下，特征得到了复制，产生了新的人群分布。在这里，均衡被定义为特征频率分布达到稳定的状态。

假设在一个两人博弈中，群体成员被随机配对，然后进行互动，他们的收益记为 $\pi(i, j)$，表示 i 特征行为者和 j 特征行为者的收益之比。因此，一个人遇到 x 型行为者的概率是 p_x，遇到 y 型行为者的概率是（$1-p_x$）。期望收益由下面的式子给出：

$$b_x(p_x) = p_x\pi(x, x) + (1-p_x)\pi(x, y)$$

$$b_y(p_x) = p_x\pi(y, x) + (1-p_x)\pi(y, y) \qquad (2)$$

第一个方程可以这样理解，x 型行为者与另一个 x 型行为者的配对概率是 p_x，他们的收益用 $\pi(x, x)$ 表示，x 型行为者遇到 y 型行为者的概率是（$1-p_x$），他们的收益用 $\pi(x, y)$ 表示。

假设在每个周期的最后，每个行为者 A 会把他这种“类型”的收益和一个随机选择的行为者 B 进行比较，比较的概率为 $\gamma_1 > 0$。如果 B 的收益比 A 低，我们就认为 A 不会改变他的文化特征。但是如果 B 的收益比 A 高，并且 B 和 A 的类型不一样，那么 A 转化为 B 的概率与 A、B 收益的差别成比例，这个比例因子是 $\gamma_2 > 0$。因此，我们可以表示如下（Gintis，1997）：

$$r_x = \gamma_1\gamma_2\ [b_x(p_x) - \bar{b}(p_x)] \qquad (3)$$

$\bar{b}$（p_x）是人群的平均收益：

$$\bar{b}(p_x) = p_x b_x(p_x) + (1 - p_x) b_y(p_x) \tag{4}$$

很明显，当且仅当 $\gamma_x = 0$ 时，人群分布 p_x 才会发生变化。我们把（3）重新表述如下：

$$r_x = \gamma_1 \gamma_2 (1 - p_x) [b_x(p_x) - b_y(p_x)] \tag{5}$$

我们发现，当且仅当满足

$$b_y(p_x) = b_x(p_x) \tag{6}$$

的时候，人群分布才会处于均衡状态。

因此，内部均衡的条件是收益相等。假设（6）的解是 p_x^*，为了保持稳定状态，p_x 的微小增加必须使得 y 特征的复制倾向大于 x 特征的复制倾向，从而增加 y 的复制量、降低 p_x。可以表示如下，

$$\frac{\mathrm{d}r_x}{\mathrm{d}p_x} < 0 \tag{7}$$

必要条件是

$$\pi(y, x) - \pi(x, x) > \pi(y, y) - \pi(x, y) \tag{8}$$

现在，我们开始来分别讨论共同体治理中的四个效应。

4 声　誉

假设每个行为者都属于两种行为者类型中的一种，我们称之为“正派的”和“下流的”。行为者通过收益 $\delta > 0$ 的观察成本来判断出哪一个行为者是“正派的”。[12] 一个正派的行为者是指，要么他可以无条件的合作，要么他对另外的正派行为者采取合作的态度，同时对下流的行为者采取背叛的态度。除此之外，其余的行为者都是下流的。在图 4 中，我们列出了 6 种纯粹策略。我们只给其中的 3 种策略进行了命名，因为其他策略是严格的占劣策略，并不会在纳什均衡中出现：(a) 和

(c) 是由于背叛而严格占劣的，而 (b) 是由于信任而占劣的。

策略	观察	行为	频率
背叛	不进行	背叛	$1-\alpha-\beta$
诚实	不进行	合作	β
(a)	进行	背叛	—
(b)	进行	合作	—
观察	进行	如果行为者下流则背叛	α
		如果行为者正派则合作	—
(c)	进行	如果行为者正派则背叛	—

图 4　囚徒困境中存在观察变量时的策略

图 5 表示愿意进行互动的一对行为者的收益矩阵。如果所有的行为者都是背叛的，我们就把纳什均衡称为普遍背叛均衡（universal defect equilibrium），如一些行为者进行了观察但是没有行为者相信，我们称这种均衡为不信任均衡（nontrust equilibrium），如果至少存在一个行为者相信这种观察，我们就称这种均衡为信任均衡（trust equilibrium）。除此之外，不再存在其他的均衡。[13] 毫无疑问，普遍背叛均衡存在，并且从逻辑上讲，它是稳定的，因此，背叛是一种演化稳定策略。

	观察	诚实	背叛
观察	$b-\delta$，$b-\delta$	$b-\delta$，b	$c-\delta$，c
诚实	b，$b-\delta$	b，b	d，a
背叛	c，$c-\delta$	a，d	c，c

图 5　观察变量存在时的囚徒困境收益矩阵

为了研究均衡存在的概率，我们用 $\alpha \geqslant 0$，$\beta > 0$，$(1-\alpha-\beta) \geqslant 0$ 分别表示进行策略观察、诚实和背叛的概率。如果不存在背叛，那么，观察相对于诚实而言将会是占劣策略，因为观察者付出了成本但没有找到背叛者，这样所有的行为者都将会是诚实的。但是接着背叛就会占优于诚实，这是个矛盾。因此，存在一个正水平的背叛行为。如

果此时没有观察行为，那么背叛就会又一次占优诚实，均衡也就无法实现。 因此，在均衡中，如果存在诚实者（如 $\beta > 0$），那么三种策略都将会是正水平的。

现在，我们来确定在一个诚实均衡中诚实、观察和背叛这三种人群的分布。 设π^i（α， β）是采用 i 策略时的期望收益。 然后，根据 (6)，在均衡中，每种行为的收益都要相等。 因此，我们得出：

$$\pi^i(\alpha, \beta) = \pi^T(\alpha, \beta) = \pi^D(\alpha, \beta),$$

或者

$$\alpha(b-\delta) + \beta(b-\delta) + (1-\alpha-\beta)(c-\delta)$$

$$= (\alpha+\beta)b + (1-\alpha-\beta)d \tag{9}$$

$$= \alpha c + \beta a + (1-\alpha-\beta)c \tag{10}$$

式子 (10) 意味着（用星号表示均衡值）：

$$\alpha^* + \beta^* = 1 - \frac{\delta}{c-d} \tag{11}$$

从这个式子可以清楚地看出，趋社会策略（观察或者诚实）与信息成本 δ 成反比，当 $\delta = 0$ 时，α、β 之和为 1。 更进一步，我们来解 (9) 和 (10)，可以得到：

$$\alpha^* = \frac{1}{a-c}\left[(a-b)\left(1-\frac{\delta}{c-d}\right) + \delta\right] \tag{12}$$

$$\beta^* = \frac{1}{a-c}\left[(b-c) - (b-d)\frac{\delta}{c-d}\right] \tag{13}$$

这样的解决方案要存在，要求α^*， β^*都大于 0，式子 (11) 表明必须 $\delta < c-d$。为了保证 $\beta^* > 0$，从式子 (13) 中我们可以知道

$$\delta < (c-d)\frac{b-c}{b-d} \tag{14}$$

注意，式子 (14) 的右边是严格正的，即有 $\delta > 0$， $\alpha > 0$。因为式子 (14) 还包含着 $\delta < c-d$，同时，考虑到囚徒困境收益中的结构

(1)，我们可以发现 (14) 是存在正水平诚实的纳什均衡混合策略的充分必要条件。因此背叛 $1-\alpha^*-\beta^*$ 的频率是 $\delta/(c-d)$，这是一个观察成本的递增函数。我们还发现诚实 β^* 的频率是 δ 的递减函数，因为从 (13) 中，我们可以得到：

$$\frac{d\beta}{d\delta}=-\frac{b-d}{(c-d)(a-c)}<0 \qquad (15)$$

因为均衡时所有策略的收益相等，即所有策略的收益和诚实的收益相等，从式子 (9)，我们知道期望收益是 $(\alpha^*+\beta^*)(b-d)$。从式子 (11)，我们有：

$$d+(b-d)(\alpha^*+\beta^*)=b-\frac{b-d}{c-d}\delta \qquad (16)$$

该式子的值随着观察成本 δ 的增加而减少。[14]

总之，在共同体提供低成本信息的基础上，不同类型的行为者互动，从而产生了四种不同的声誉效应。首先，信息成本的减少使均衡得以实现，从而产生诚实行为者，如 (14)。第二，在这样的均衡中，诚实行为者的数量比较大，信息成本比较低 (15)。第三，如果诚实发生在均衡之中，那么人群的平均收益与信息成本成反向关系 (16)。最后，人群中的背叛行为将会直接随着信息成本而发生改变 (11)。

5 报　　复

如果图 2 中的囚徒困境以某种概率重复进行的话，那么由于对背叛者报复威胁的存在，合作将会得以进行，并且威胁越有效，重复的可能性也就越大。如果重复次数足够，并且每次重复的间隔时间很短，那么收益结构就会发生改变，从而出现两种均衡：普遍背叛 (universal defect) 和普遍合作 (universal cooperate)。

改变后的博弈成为信任（assurance）博弈（Sen，1967），因为在合作中，每个行为者都会做到最好，只要能够确信其他的合作也能做到最好。[15]

在原来的囚徒困境中，背叛是优势策略（即无论另外的行为者采取什么样的行动，都能给行为者比较好的收益）。但在信任博弈中，每个行为者都相信别人会采取合作策略，因此社会最优产出（互相合作）是均衡状态。我们将会看到，共同体的高额进入和退出成本和共同体成员之间的重复互动，可能会把一个难以处理的合作问题转变为一个更经得起检验的解决方案。我们还会看到，共同体会增加合作均衡的吸引域，因此即使在随机干扰的情况下，合作的结果也会更加稳健。

由于共同体具有高额的进入和退出成本，个体间互动就更加频繁和具有重复性，这就使得在短暂条件下不能实现的合作结果得以实现。

重复通过两种方式改变互动，更复杂的策略产生了。一种是考虑到行为者以前的行为，这就需要把收益看成是整个时期内的期望所得。于是，行为者就会采用所谓的以牙还牙（Tit-for-Tat）策略：在第一个回合中采取合作策略，随后的策略则取决于对手前一回合所采取的策略。为了简单起见，我们把策略限制在以牙还牙（T）和无条件背叛（D）之间。现在我们可以来计算期望收益了。

假设在每个回合之后，互动终止的概率是 ρ，并且每次重复的时间间隔很短，因此可以忽略时间偏好率。例如，当两个 T 型行为者相遇时，他们会采取合作的方式直到互动终止（期望的持续时间是 $1/\rho$），期望收益是 b/ρ。当一个 T 型行为者遇到一个 D 型者，前者在第一回合中将会得到收益 d，然后双方都将采取背叛的方式直到博弈终止，第一回合之后的期望博弈数是 $1/\rho-1=(1-\rho)/\rho$，期望收益是 $d+(1-\rho)c/\rho$。详细的收益矩阵见图 6。

如果人群中采用以牙还牙的部分是 τ（其余的采取无条件背叛），并且博弈的双方是随机配对的，则和一个以牙还牙者配对的概率是 τ，T 和 D 型行为者的期望收益分别用 π^T、π^D 表示：

	以牙还牙	无条件背叛
以牙还牙	b/ρ b/ρ	$d+(1-\rho)c/\rho$ $d+(1-\rho)c/\rho$
无条件背叛	$a+(1-\rho)c/\rho$ $d+(1-\rho)c/\rho$	c/ρ c/ρ

图 6　重复囚徒困境博弈的收益矩阵(ρ 是博弈终止的概率)

$$\pi^T(\tau)=\tau b/\rho+(1-\tau)\{d+(1-\rho)c/\rho\} \qquad (17)$$

$$\pi^D(\tau)=\tau\{a+(1-\rho)c/\rho\}+(1-\tau)c/\rho \qquad (18)$$

均衡时，τ^* 为：

$$\tau^*=\frac{c-d}{2c-a-d+(b-c)/\rho} \qquad (19)$$

对终止概率有：

$$\rho<\frac{b-c}{a-c} \qquad (20)$$

对于 $c>d$，我们有 $\tau^*\in(0,1)$，从而给出内部均衡。注意 (20) 也保证了 (19) 的分母为正数。第二个条件必须是真的，因为对于单期博弈来说，收益矩阵是会出现囚徒困境的。当普遍协调的收益比单期背叛的收益大的时候，第一个条件是正确的。上面所提到的收益和内部均衡τ^*可以用图 7 来表示。

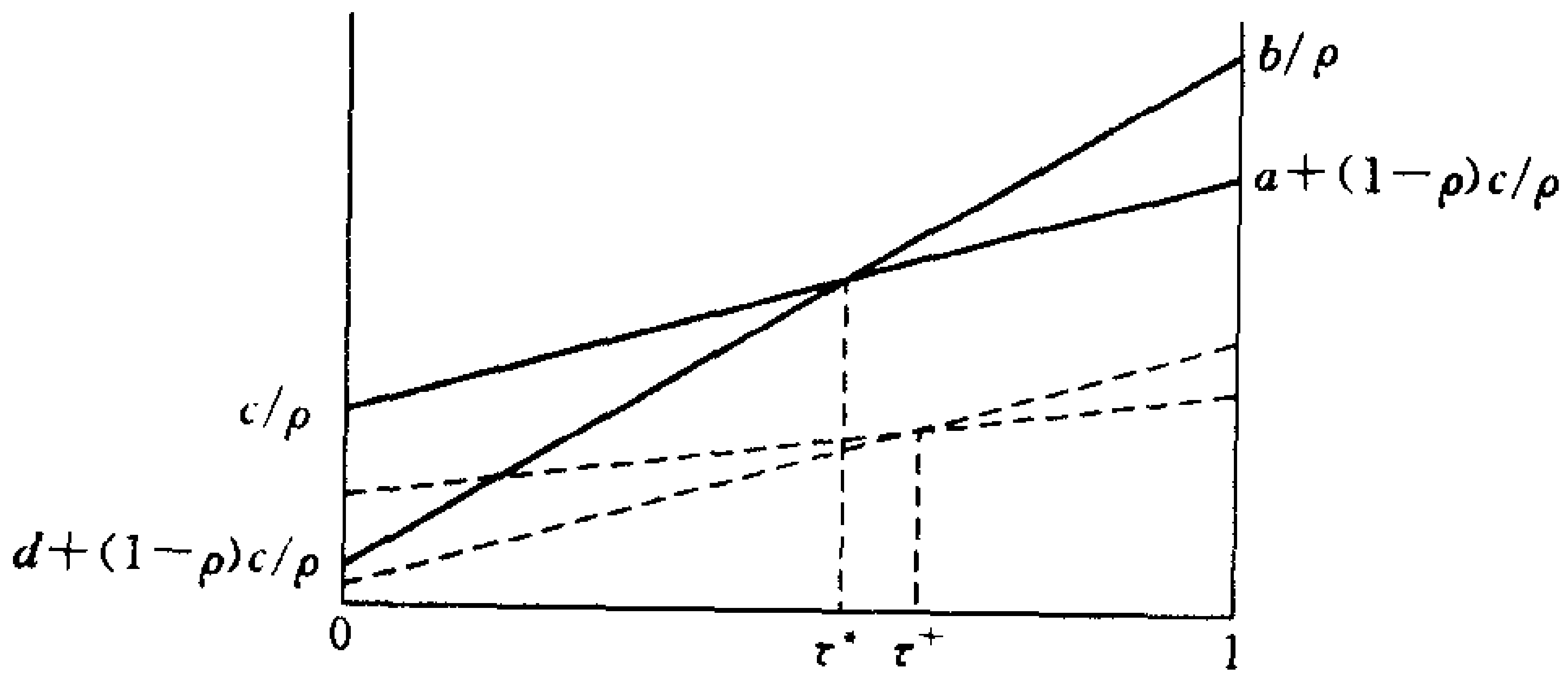

图 7　报复效应(终止概率(虚线)的增加会减少，甚至消除普遍合作均衡的吸引域)

与声誉效应的博弈不同，τ^* 是不稳定的，τ^* 的微小偏离不会回到其收敛状态。这是因为有：

$$\frac{\mathrm{d}\pi^D(\tau^*)}{\mathrm{d}\tau} < \frac{\mathrm{d}\pi^T(\tau^*)}{\mathrm{d}\tau}$$

这个条件与稳定性条件（7）相违背。我们可以从下面的考察来探讨这一点。当 $\tau > \tau^*$ 的时候，D 的期望收益相对于 T 来讲是递减的，但是，如果收益等于 τ^*，D 的收益就会比 T 小，在本文的第三部分，我们描述了动态过程，在这样的过程里，对 T 而不是 τ^* 的回报增加了。因此有三种均衡的人群频率，分别为 0，τ^* 和 1。第一和第三是稳定的，不稳定均衡 τ^* 描述了两种稳定均衡的分界状态。

式子（20）已经表明，当人群中没有背叛者的时候，以牙还牙的收益会超过人群中存在背叛行为时的收益，因此以牙还牙是一种比较好的反应策略。如果人群中背叛者的人数比例是正数 ε，我们就称以牙还牙是一种演化稳定策略，因此，如果背叛的人数比例小于 ε，那么异质复制的过程会导致背叛行为消失（Weibull，1995）。当背叛者与总人口的比例小于 ε 时，入侵就会失败。因此，以牙还牙是一种演化稳定策略，根据上面的定义，临界值 ε 等于 $1-\tau^*$。

共同体的治理效应产生了两个结果。首先，如果终止的概率足够低，那么互动就会产生一个普遍合作的均衡（普遍背叛会保持一种均衡）。这可以从（20）直接得出：如果内部均衡存在并且是不稳定的，那么 $\tau=0$ 是一种必需的稳定均衡。第二，终止概率的增加会减少合作均衡的吸引域。这是因为：

$$\frac{\mathrm{d}\tau^*}{\mathrm{d}\rho} = \frac{(\tau^*)^2(b-c)}{\rho^2(c-d)}$$

如果初始收益是囚徒困境式的，并且 $\tau^* \in (0,1)$ 的话，那么上式就是正的。随着互动持续时间的减少（ρ 增加），普遍背叛和普遍合作的程度分界线会朝着后者移动，扩大了初始条件的范围，从而使得 $\tau^*=0$。

条件（20）并不能保证普遍合作一定发生。它只能保证普遍合作一旦发生，那么单方的背叛过程并不能拆开这种合作。这是因为共同体交往的连续性（较小的 ρ）使合作成为可能。

6 分　割

共同体的高额进入和退出成本使得人群被分割，共同体内部成员之间的交往，远远频繁于内部成员和外部成员间的交往。例如，小村庄的人们内部互动非常频繁，他们只是偶尔在一个单独的市场上和其他人群交往。

假设在一个单期的囚徒困境博弈中，个人要么是背叛者要么是合作者，他们根据两种策略的相对成功性来定期重新确立他们的类型。与声誉和报复模型相反，分割模型是建立在非随机配对基础之上的。共同体的人们被分割成同质性更强的人群，这种分割要么是由于血缘关系的亲近，要么是基于文化特征的相关性。[16]

"物以类聚、人以群分"减少了合作问题的产生，诸如囚徒困境中的趋社会行为者会采取合作的手段，把利益给了与自己互动的人，那是因为背叛是会产生成本的。因此，一个有偏见的（biased）配对过程会提高趋社会行为者的收益。与共同体有关的分割使得趋社会行为者可以获得较多的收益，因此这种特征在人群中就会得到传播。

我们用下面的方式来定义分割的程度。如果人群中合作者的比例是 α，这样一个协调者遇到另一个协调者的概率不再是 α，而是 $\sigma+(1-\sigma)\alpha$，$\sigma\in(0,1)$，σ 代表人群分割程度。相应的，一个背叛者遇到另一个背叛者的概率是 $\sigma+(1-\alpha)(1-\sigma)$。注意当 $\sigma=1$ 时，表示有共同爱好的人进行配对，当 $\sigma=0$ 时表示配对是随机的。因此从 Hamilton 定律来说，分割程度和相关性是等同的，控制着利他行为者的演化（Grafen, 1979; Grafen, 1984;

Axelrod and Hamilton, 1981)。如果整个族群都是同质的，那么就会出现一种特别简单的情况，σ 表示从族群中选择互动对象的概率，而不是表示从整个人群中选择互动对象的概率。拿上面的例子来说，就是村庄内的人们进行交易，而不是在整个大市场内进行交易。期望收益可以表示成：

$$\pi^C(\alpha, \sigma) = \sigma b + (1-\sigma)[\alpha b + (1-\alpha)d]$$

$$\pi^D(\alpha, \sigma) = \sigma c + (1-\sigma)[\alpha a + (1-\alpha)c]$$

考虑到共同体的聚集特征，我们把配对规则和分割程度当成共同体所支持的聚类 (clustering of types) 的给定外在因素，现在我们来考虑它在合作的均衡水平中的作用。[17] 为了研究这种作用，我们发现，α 的数值等于上面两个期望收益，或者

$$\alpha^* = \frac{\sigma(d-b)+c-d}{(1-\sigma)(b-d-a+c)}$$

这个均衡是否稳定取决于收益水平。在后面的例子里，α^* 表示 $\alpha=1$ 和 $\alpha=0$ 两种稳定均衡的吸引域之间的临界线。图 8 表示的是稳定内部均衡的例子。[18]

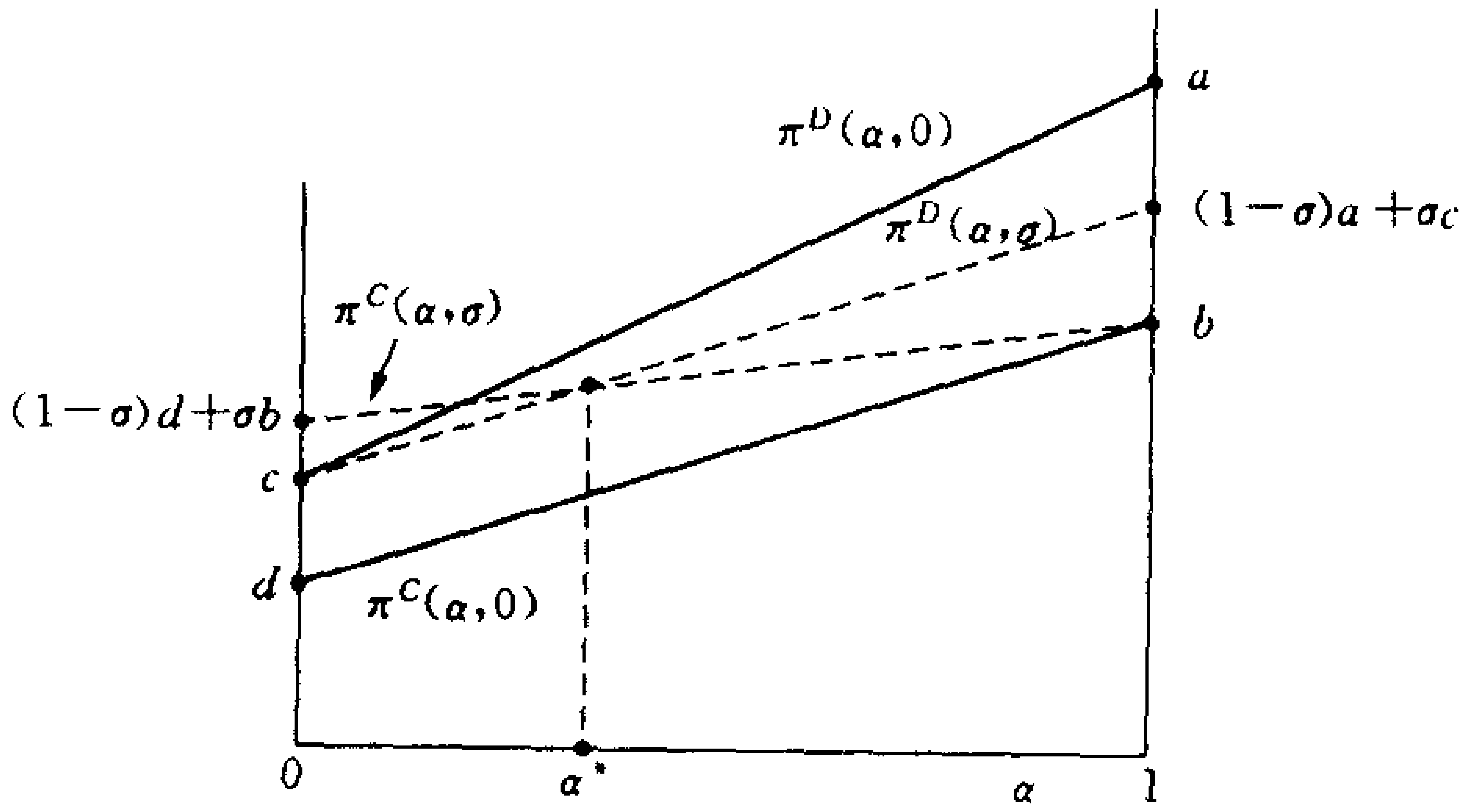

图 8 稳定内部均衡时的分割效应

四个结论支持了我们对分割效应的解释。首先，存在着某个 $\sigma<1$，因此普遍合作是一种均衡状态，即使行为者之间的交往是一种单期的

囚徒困境博弈。我们把这个分割程度的临界值记为 σ'，此时 $\alpha^* = 1$。因此有：

$$\sigma' = \frac{a-b}{a-c} < 1$$

这个不等式成立的原因是，囚徒困境下的收益指定 $b > c$。

第二，存在某个数值 $\sigma < 1$，称为 σ''，当 $\sigma > \sigma''$ 时，某种程度的合作可以像均衡那样维持，这时 $\alpha = 0$，或者

$$\sigma'' = \frac{c-d}{b-d}$$

因为 $c < b$，所以上式小于 1。

第三，如果 α^* 是稳定的，那么分割的增加会导致合作人数的增加。因为 $d\alpha^*/d\sigma$ 和 $(c-b)(b-d-a+c)$ 的正负号相同，在一个稳定均衡中，这是个正数。

第四，如果 α^* 是不稳定的，那么 α^* 将区分出所有的背叛稳定均衡和所有的合作稳定均衡，分割的增加会扩大普遍合作均衡的吸引域，因为在这里，有 $d\alpha^*/d\sigma < 0$。

7 狭隘的地方观念

如果一个人群中的子族群表现出不同层次的趋社会规范，并因此而遭遇了不同程度的合作失败，那么，大量迁徙到相对趋社会的族群中也许会使得合作均衡难以达成。“地方”文化价值观降低了迁徙率，这样与趋社会规范本身的交互能够帮助共同体维持稳定的合作互动。为了说明这一点，我们引用了 Boyd 和 Richerson (1990)的囚徒困境模型。

我们先回到报复效应模型，但现在把族群加入到第五部分所提到的模型中。互动只在族群内部发生，但在每个时期中，族群之间会出现迁移，假设每个周期的迁移数占族群人数的比例是 μ。

我们现在来讨论迁移过程。如前所述，个体以终止概率 ρ 在一定时期内互动，并且在交往终止后，他们通过观察其他人的收益来更新他们的行为。在这种情况下，族群中的 $\mu \in (0, 1)$ 部分的人迁移出去，而同等比例的外部成员则迁移进来。关于迁移的更加复杂和现实的模型——迁移的人会把成功的族群作为选择对象——不会改变（Bowles and Gintis, 1997）结果。进入和退出的成本越高，μ 就越低。

假设那些在一个特定族群中运用以牙还牙策略的比例是 τ，这个比例由于行为的更新而随时间发生改变，有：

$$\tau' = \tau + \dot{\tau}\, dt$$

迁移改变了人口组成，我们根据

$$\tau'' = (1 - \mu\, dt)\,\tau' + \mu\,\bar{\tau}\, dt$$

把迁移前的频率转换为迁移后的频率 τ''。在这里，$\bar{\tau}$ 表示总人口中以牙还牙行为者的比例。[19]

族群中以牙还牙行为者的均衡频率必须满足 $\tau = \tau''$（频率必须是固定的），或者满足，

$$\frac{\dot{\tau}}{\tau} = \frac{\mu}{1-\mu}\left(1 - \frac{\bar{\tau}}{\tau}\right) \tag{21}$$

我们可以这样来理解（21），由于更新（式子左边）而带来的特征转换效应必须刚好被迁移所带来的效应所弥补。正如某些人所预测的，族群特征的频率等于人口的平均数，迁移对族群的频率不产生影响，因此式子（21）必须要求有 $\dot{\tau} = 0$，或者条件（6）成立。

从（5）我们知道，人口频数的增长率 $\dot{\tau}/\tau$ 可以表示如下：

$$\frac{\dot{\tau}}{\tau} = \gamma_1\gamma_2\,[\pi^T(\tau) - \bar{\pi}(\tau)]$$

$$= \gamma_1\gamma_2(1-\tau)[\pi^T(\tau) - \pi^D(\tau)]$$

应用报复博弈的收益（17）和（18），我们又可以得到下面的式子：

$$\frac{\dot{\tau}}{\tau} = \gamma_1 \gamma_2 (1 - \tau)\left[\tau \left\{2c - a - d + \frac{b - c}{\rho}\right\} - c + d\right]$$

我们运用这个表达式和上面的均衡条件来定义均衡的人口频率 τ_μ，在图 9 中，我们可以发现 $\tau_\mu > \bar{\tau}$。为了检查 τ_μ 是否稳定，我们假设 $\tau > \tau_\mu$。人口组成的迁移效应比前面所提到的由于收益变动所带来的行为更新效应要来得大，因此，$d\tau/dt < 0$。如果 $\tau < \tau_\mu$，则出现相反的情况，因此根据（7），τ_μ 是稳定的。

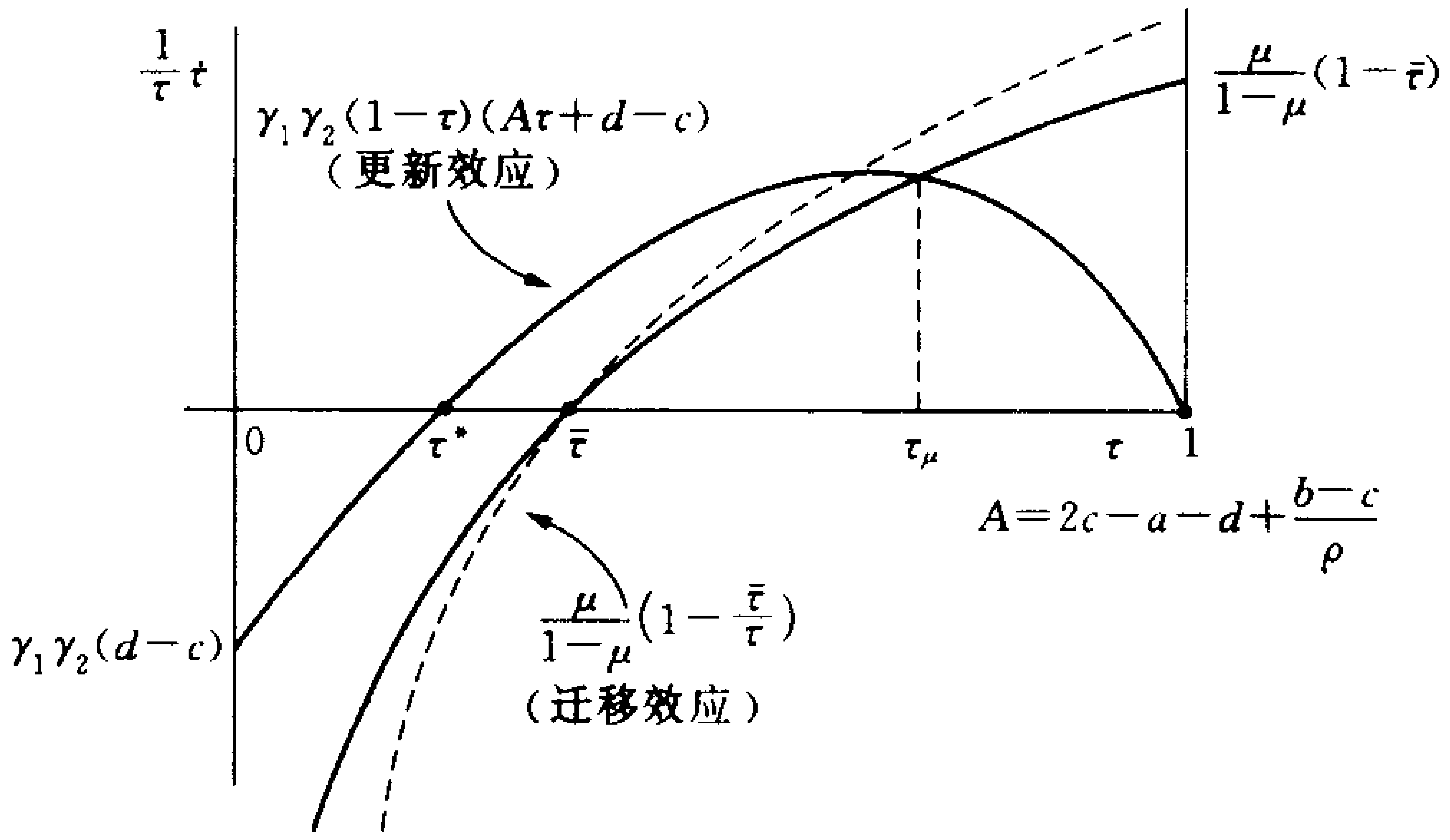

图 9　迁移下的报复效应：迁移的增加（虚线）减小了合作行为的均衡频率

在前面的报复模型中，如果终止概率足够低，则普遍合作（通过使用以牙还牙策略）是一种稳定均衡。族群成员的收益越高，则特征的繁衍就越快，这是因为特征在不同的族群中得到异质复制。族群之间的迁移以下面的方式来改变结果。

首先，如图 9 所示，如果以牙还牙的行为者数目 τ^* 比整个人群的比例 $\bar{\tau}$ 小的话，那么有：

$$\frac{d\tau_\mu}{d\mu} < 0$$

这就意味着迁移比率的增加，会减少均衡时行为者以牙还牙的频率，同时增加了均衡中行为者背叛的频率。如图 9 的虚线所示。

第二，如果 $\bar{\tau} < \tau^*$（图 9 中没有画出这种情况），那么就可能存在一种低水平合作的均衡（低于 τ^*），也可能存在三种均衡，其中两种是稳定均衡（一种合作程度比较高，另一种则比较低），一种是介于上面两种均衡之间的非稳定均衡。[20] 在这个例子中，如果原来是比较高的稳定均衡，那么迁移比率的增加会减少合作水平；如果原来是比较低的均衡水平，那么迁移比率的增加会增加合作水平。同时，存在着某个足够高的迁移比率可以消除比较高的合作均衡。

8 结　论

行为者之间的个人交往是由共同体、市场、国家、家庭和其他的制度来组织的。在这种治理的连接关系中，共同体的重要性至少在部分上反映了共同体效应所产生的收益与替代的制度结构的收益之间的平衡。尽管在这里不用模型来模拟这种过程，但我们还是有理由认为，由共同体均衡和其他治理结构所控制的人群互动，可以成功地解决协调失败，这样就会增加和占领新领域，吸收其他的人群，并因此而替代其他比较不成功的治理结构。在这种情况下，选择性的压力也许会造成军事和经济竞争，以及人们可能通过观察其他社会中的成功治理结构来取代不成功的治理结构。

尽管共同体不能充分运用市场的效率改进和国家所提供的普遍强制性规则，但是，共同体的工具使得它们可以在市场交换和现代化的国家中持续发展下去。在共同体的这些工具中，本文所研究的一个，就是共同体可以培养成员的合作行为，从而以较低的成本来解决囚徒困境的协调问题。相似的结论可以用于一般化的收益结构中（Bowles，1996）。通过诱导趋社会行为，共同体可以支持使行为变得有秩序和公正的规范和价值观，这是因为，人们总是在寻找行动和评价之间的一致性。

我们并不是说，共同体因为这些理由而存在。我们只是证明了这种存在的可能性。也有人提出其他的原因，在这些提法之中，最突出的观点是，共同体和它的价值观由于信奉者和文化传播过程的内在倾向而得以延续。我们并不怀疑这种倾向的存在，并且有时这种倾向还是决定性的。但由于本文开始所提出的那些理由，我们并不认为在解释以共同体为基础的社会互动或与此相关的社会规范时，纯粹惯性和追溯过去的方法可以提供足够的解释力。

国家和市场作为现代治理结构，其繁衍、分散、衰落和消亡受到时代进程的规制。与以前的观点不同，我们认为在未来的日子里，共同体将会在治理结构的关联性方面起重要的作用，因为共同体可以成功地解决某些市场或国家无法解决的问题。

很多人认为，随着产品由货物向服务转化，与信息相关的服务的增加（Quah，1996），以及团队生产方式的重要性的增加，合作所带来的收益也会得到增加。原因在于，监视这样的行为需要高额成本，或者说不可能实现，因此，不管是完整契约所需要的运行良好的市场，还是集权化管理所需要的良好的国家监管，都是不切实际的。在这种情况下，我们希望共同体的生存能力会逐渐增加。另外，与以共同体为基础的社会互动相联系的社会有时会出现强烈违反普遍性规范的事，也可能激发要么法律禁止要么演化上有残缺的问题，这些都不在本模型的考虑之内。

注释：

[1] E. O. Wilson (1975) 把族群定义为“任何组织体的集合，这些组织体同属于一种类，它们在一定时期内相互交往，保持一致的程度远高于和其他种类保持一致的程度。” (p. 585)

[2] 经济学家和政治学家对“俱乐部”进行了多方面的研究，但是“俱乐部”和我们的“共同体”有不同的地方，俱乐部有一个常规的决策机构，并且提供公共产品。

[3] 当然，对于我们所定义的共同体，它们也有可能妨碍效率的提高。

[4] 文化的传播过程可以是“传教式”的，因此，人群中的个体就以较高的频率模仿文化形态。文化演化的关键文献是 Cavalli-Sforza and Feldman(1981)和 Boyd and Richerson(1985)。

[5] 我们用信息理论来解释市场、国家和共同体的优点和缺点。详细内容见：

Bowles and Gintis(1998), Farrell(1987)。

[6] 第二个条件排除了背叛者和协调者交换所带来的社会最优性。只要我们有 $a+d>2c$，那么协调就是普遍的社会选择，尽管我们不需要这个事实。

[7] 注意，在盲从的文化传播缺位的情况下，我们采用了趋社会特征来支持这种均衡。我们可以认为，地方观念和族群选择都可能对趋社会特征的演化有贡献（Wilson, 1980; Boyd and Richerson, 1990; Soltis, Boyd and Richerson, 1995; Wilson and Sober, 1994）。在这些力量无效的情况下，我们用模型来对此进行研究。我们的本位主义模型是由 Hamilton(1975)提供的，他认为，利他主义的特征在单个族群中会得到扩散，这主要是因为没有外来人口的迁入，导致了遗传相关性的增加。

[8] 这个框架是根据注 4 所提及的文献的得来的。还有其他定义文化的方式，但它们所强调的方面，如文化的功能性和法律性角色，以及文化在历史传统中的综合性质和基础作用等等，在这里，我们都把它们排除在外。

[9] 当然，在文化传播结构下，人类社会所需要的文化特征是文化本身所带来的遗传和文化演化过程。

[10] 这个框架是从前面的文化演化模型中得来的，但它也和其他的方法相一致。例如，可以参见 Bandura(1977)。

[11] 我们假设人口数量足够大，这样我们就可以把 γ_x 和 p_x 当作实际的数字。

[12] 对观察和诚实的这种处理是根据 Güth and Kliemt(1994)。这个动态模型是鲍尔斯和金迪斯（1997）全面发展出来的。

[13] 我们将会假设 x 型的行为者足够多，这样，我们就可以把 x 作为一个连续的真实变量。特别的，我们假设所有的行为者都选择了适当的纯战略以满足纳什均衡，我们也允许 x 的函数在正实数上是连续的。

[14] 为了更加全面地进行讨论，我们必须处理当只存在观察和背叛两种行为时的情况。如果观察的概率是 $\alpha>0$，那么均衡时背叛的概率是 $1-\alpha^*>0$。因为背叛的收益是 c，观察的收益也一定是 c。因此在普遍背叛均衡中，这样的一种均衡并不能产生社会效益。这种均衡是动态不稳定的，因此在现实中就不能观察到（Bowles and Gintis, 1997）。我们将不再对这种均衡作进一步的讨论。

[15] 该博弈也被称之为“会猎（stag hunt）”，是 J.-J. Rousseau（1755/1987）在他的寓言中给出的。

[16] 例如，如果某一群体是具有这些特征的人的后代，并且从父母那里得来的直接遗传很明显的话，那么这个群体的同质性就比较强。

[17] 在族群选择的压力下，配对规则和 σ 也许会发生改变，尽管不存在必要的例子。我们在这里不讨论这种可能性。

[18] 第三部分给出的稳定条件是上式的分母为负数，且 $\alpha>0$，分子也为负数。我们可以从图 8 中的纵轴截距很明显地看出这一点。当单方面背叛（$a-b$）的收益比惩罚的损失（$d-c$）大的时候，我们可以取得一个稳定的均衡。

[19] 为研究的简便，我们假设，相对于共同体来讲，总人口足够多，则 τ 不受迁移的影响。

[20] 这可能和均衡同时发生。

参考文献：

Axelrod, Robert, *The Evolution of Cooperation* (New York: Basic Books, 1984).

— and William D. Hamilton, “The Evolution of Cooperation,” *Science* 211 (1981):1390—1396.

Bandura, Albert, *Social Learning Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1977).

Bowles, Samuel, “Markets as Cultural Institutions: Equilibrium Norms in Competitive Economies,” 1996. University of Massachusetts Discussion Paper.

— and Herbert Gintis, *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life* (New York: Basic Books, 1976).

— and —, “Optimal Parochialism: The Dynamics of Trust and Exclusion in Communities,” June 1997. University of Massachusetts Working Paper.

— and —, *Recasting Egalitarianism: New Rules for Markets, States, and Commu-*

nities (London: Verso, 1998). Erik Olin Wright (ed.).

Boyd, Robert and Peter J. Richerson, *Culture and the Evolutionary Process* (Chicago: University of Chicago Press, 1985).

— and —, "Group Selection among Alternative Evolutionarily Stable Strategies," *Journal of Theoretical Biology* 145 (1990):331—342.

Boyer, Paul and Stephen Nissenbaum, *Salem Possessed: The Social Origins of Witchcraft* (Cambridge: Harvard University Press, 1974).

Cavalli-Sforza, Luigi L. and Marcus W. Feldman, *Cultural Transmission and Evolution* (Princeton: Princeton University Press, 1981).

Farrell, Joseph, "Information and the Coase Theorem," *Journal of Economic Perspectives* 1, 2 (Fall 1987):112—129.

Festinger, Leon, *A Theory of Cognitive Dissonance* (Stanford: Stanford University Press, 1957).

Flannery, Kent, Joyce Marcus, and Robert Reynolds, *The Flocks of the Wamani: A Study of Llama Herders on the Puntas of Ayacucho, Peru* (San Diego: Academic Press, 1989).

Fromm, Erich and Michael Maccoby, *Social Character in a Mexican Village: A Sociopsychoanalytic Study* (Englewood Cliffs: Prentice-Hall, 1970).

Fudenberg, Drew and Eric Maskin, "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica* 54, 3 (May 1986): 533—554.

Geertz, Clifford, *Peddlers and Princes: Social Change and Economic Mod-ernization in Two Indonesian Towns* (Chicago: University of Chicago Press, 1963).

Gintis, Herbert, "Welfare Economics and Individual Development: A Reply to Talcott Parsons," *Quarterly Journal of Economics* 89, 2 (June 1975).

—, "The Power to Switch: On the Political Economy of Consumer Sovereignty," in Samuel Bowles, Richard C. Edwards, and William G. Shepherd (eds.) *Unconventional Wisdom: Essays in Honor of John Kenneth Galbraith* (New York: Houghton-Miffin, 1989) pp.65—80.

—, A Markov Model of Production, Trade, and Money: Theory and Articial Life imulation, *Computational and Mathematical Organization Theory* 3, 1 (1997):19—41.

Grafen, Alan, "The Hawk-Dove Game Played between Relatives," *Animal Behavior* 27, 3 (1979):905—907.

—, Natural Selection, Kin Selection, and Group Selection," in J. R. Krebs and N. B. Davies (eds.) *Behavioural Ecology: An evolutionary Approach* (Sunderland, MA: Sinauer, 1984).

Güth, Werner and Harmutt Kliemt, "Competition or Co-operation: on the Evolutionary Economics of Trust, Exploitation, and Moral Attitudes," *Metroeconomica* 45, 2 (1994):155—187.

Hamilton, W. D., "Innate Social Aptitudes of Man: an Approach from Evolutionary Genetics," in Robin Fox (ed.) *Biosocial Anthropology* (New York: John Wiley and Sons, 1975) pp. 115—132.

Harding, Susan, "Street Shouting and Shunning, : Conict between Women in a Spanish village," *Frontiers* Ⅲ, 3 (1978):14—18.

Jenness, Diamond, *Arctic Odyssey: the Diary of Diamond Jenness, Ethnologist with the Canadian Arctic Expedition in Northern Alaska and Canada*, 1913—1916 (Hull, Quebec: Canadian Museum of Civilization, 1991).

Kelly, Raymond C., The Nuer Conquest: *The Structure and Development of an Expansionist System* (Ann Arbor: University of Michigan Press, 1985).

Kohn, Melvin, *Class and Conformity* (Homewood, IL: Dorsey Press, 1969).

Kreps, David M., "Corporate Culture and Economic Theory," in James Alt and Kenneth Shepsle (eds.) *Perspectives on Positive Political Economy* (Cambridge: Cambridge University Press, 1990) pp. 90—143.

LeVine, Robert A., *Dreams and Deeds*: Achievement Motivation in Nigeria (Chicago: University of Chicago Press, 1966).

Light, Ivan, Im Jung Kwuon and Deng Zhong, "Korean Rotating Credit Associations in Los Angeles," *Amerasia* 16, 1 (1990):35—54.

Malinowski, Bronislaw, *Crime and Custom in Savage Society* (London: Routledge & Kegan Paul, 1926).

Platteau, Jean-Philippe, "Traditional Sharing Norms as an Obstacle to Economic Growth in Tribal Societies," *Cahiers de la Faculté des Sciences Economiques et Sociales, Facultes Universitaires Notre-Dame de la Paix* 173 (1996):201—223.

Quah, D., "The Invisible Hand and the Weightless Economy," 1996. Centre for Economic Performance, London School of Economics.

Rousseau, Jean-Jacques, "Discourse on the Origin and Foundations of Inequality Among Men," in Donald A. Cress (ed.) *Basic Political Writings* (Indianapolis: Hackett Publishing Company, 1987) pp. 25—109.

Sen, Amartya K., "Isolation, Assurance, and the Social Rate of Discount," *Quarterly Journal of Economics* 81 (1967):1112—1124.

Shapiro, Carl, "Premiums for High Quality Products as Returns to Reputations," *Quarterly Journal of Economics* (1983):659— 679.

Soltis, Joseph, Robert Boyd and Peter Richerson, "Can Group-functional Behaviors Evolve by Cultural Group Selection: An Empirical Test," *Current Anthropology* 36, 3 (June 1995):473—483.

Taylor, Michael, *The Possibility of Cooperation* (Cambridge: Cambridge University Press, 1987).

Weber, Eugen, *Peasants into Frenchmen: The Modernization of Rural France, 1870—1914* (Stanford: Stanford University Press, 1976).

Weibull, Jörgen W., *Evolutionary Game Theory* (Cambridge: MIT Press, 1995).

Wilson, David Sloan, *The Natural Selection of Populations and Communities* (Menlo Park, CA: Benjamin Cummings, 1980).

— and Elliott Sober, "Reintroducing Group Selection to the Human Behavioral Sciences," *Behavior and Brain Sciences* 17 (1994):585—654.

Wilson, Edward O., Sociobiology: *The New Synthesis* (Cambridge: Harvard University Press, 1975).28

私有财产的演化

赫伯特·金迪斯*

1 引　言

如果一个行为者对某物拥有排他性的使用权而且能从这个专断性的使用权中得到收益，那么我们可以说该行为者拥有该物品。进一步说，如果该所有权很少受到威胁，而一旦受到威胁，原所有者一般还能继续拥有该所有权，那么所有权就是得到尊重的（respected）。

从霍布斯、洛克、卢梭和马克思直到现在，西方思想史的主流观点总是认为私有财产是一种伴随着现代文明兴起而出现的人类社会结构（Schlatter，1973）。但是，在过去四分之一个世纪里，我们从动物行为研究中所收集到的证据已经说明了这个观点是不正确的。我们在非人类物种中发现各种各样领地宣告（territorial claim）的情况，这些物种包括蝴蝶（Davies，1978），蜘蛛（Riechert，1978），野马（Stevens，1988），雀类（Senar，Camerino and Metcalfe，1989），黄蜂（Eason，Cobbs and Trinca，1999），灵长类（Ellis，1985）和许多其他生物（Mes-

* 原文题目为 The Evolution of Private Property，unpublished working paper，谢家骏译。我要感谢 Carl Bergstorm 和 John Maynard Smith 给我的评论，他们的评论给了我很大的帮助，我还要感谢 John D. 和 Catherine T. MacAthur 基金为我提供的资助。

terton-Gibbons and Adams, 2003)。进一步的研究显示，幼儿和儿童运用那些类似于动物的行为规则来认可和保护产权（Furby，1980）。

在非人类生物中，一群动物拥有一块领地一般是建立在这群生理上接近的动物控制领地以及打算以维持和开发的方式来转变（transform）领地的事实之上的。在人类社会中还有其他的所有权标准，但是上述的原则依然是重中之重。下面我们以约翰·洛克的说法为例：

> ……每人对他自己的人身享有一种所有权……他的身体所从事的劳动和他的双手所进行的工作，我们可以说，是正当地属于他的。所以只要他使任何东西脱离自然所提供的和那个东西所处的状态，他就已经掺进他的劳动，在这上面参加他自己所有的某些东西，因而使它成为他的财产。
>
> 《政府论》，§27(1690)

至于所有权是怎样演变的以及它在演化的背景中又是如何维持的，这是一个极富挑战性的谜题。试想一下麻雀在我家花园的葡萄藤上筑了一个巢。它们先要选择巢的位置，然后花费数天时间来准备鸟巢的构建。当然这个巢对于其他的麻雀来说也会具有同等的价值。但是为什么其他的麻雀不试图把第一对麻雀驱逐出去呢？如果双方同样强壮，而且对巢价值的评价也相同，那么双方各有一半的机会赢得这场领地争夺战的胜利。

一个常见的观点认为，最初的那一对麻雀会失去更多，因为它们已经为改善财产付出了巨大的努力。但是，这个解释却犯了被称为合成谬误（Dawkins and Brockmann, 1980）的逻辑错误：为了最大化适存度，一个行为者只会考虑一个实体未来的收益，而不考虑过去已经为该实体所付出的成本。另一个曾被认可的观点认为所有权是出于对整个群体利益的考虑，但是 20 世纪 60 年代出现的关于群体选择的批评（Maynard Smith, 1976; Williams, 1966; Dawkins, 1976）否定了

这种解释。毫无疑问，在许多环境下通过认可所有权可以提高一个群体中所有成员的适存度。但是问题在于，一个只关心自己后代利益的突变体可以非常轻易地破坏产权并使得破坏产权的基因扩散。换句话说，对所有权的尊重并不是一种明显的“演化稳定策略”（Maynard Smith and Price，1973）。

在成功地对私有财产的群体选择解释提出批评以后，John Maynard Smith 又创造性地提供了一种新的选择（Maynard Smith，1982），即现在非常有名的鹰—鸽博弈模型。在这个博弈模型里，老鹰和鸽子属于同一种类中不能通过外部表征加以区分的成员，但是在对领地的竞争中却表现得十分不同。当两只鸽子竞争领地的时候，它们会装模作样一番，然后对等地平分领地（或者说，双方各有一半的机会占有这块领地）。但是当一只鸽子和一只老鹰竞争领地时，老鹰则会拥有整块领地。最后，当两只老鹰竞争时，一场可怕的战斗将会随之而来，而且领地的价值会小于竞争双方付出的成本。Maynard Smith 指出，在种群中存在着一种有部分行为者扮演模型中老鹰的角色、剩下的扮演鸽子的演化稳定策略。然后 Maynard Smith 考虑了第三种策略：如果你是第一个发现这块领地的，就像老鹰那样行为；若不是第一个发现者，则像鸽子那么行为。我们非常容易看出这种把老鹰和鸽子都排除在外的“中庸”策略不但是演化的而且是全局稳定的。我们很难再想像出一个解答私人财产问题的更有说服力的答案。

但是鹰—鸽模型还是存在着几个问题。首先，中庸策略是全局稳定的。这意味着只要种群中的成员有能力分辨出谁先占有领地，而且有能力在出现争夺战的情况下让对手遭受到足够沉重的打击，那么私人财产的利害关系就会受到重视。而事实上，私人财产看上去比我们根据上述推断所期望的要少得多。在竞争者—侵占者模型中，私人财产均衡通常是演化稳定的，却很少是全局稳定的，所以当一小部分保护财产的突变体在面对大量不承认产权的行为者时就不能成功地侵犯。

关于全局稳定性的争议引出了私人财产概念中最为核心的部分。

在鹰—鸽模型的框架下，无论从哪一点出发，根据全局稳定性，私人财产都将成为演化过程中不可避免的终点。事实上，私人财产被看成一种一旦建立就能延续下去的惯例更为恰当。以一只标记领地和准备保卫领地的鸟为例，当大量同类侵占其中任何一部分能满足它们目标的领地时，它是无法抵挡这种冲击的。所以自然中的私有财产奇迹可以看成是一个基因编码的社会惯例。

鹰—鸽模型的第二个问题是 Grafen（1987）提出的，他注意到了领地所有权的成本和收益并不是固定的，而是依赖于种群的数量、高质量领地的密度、搜寻的成本以及其他有关种群中分配策略的变量。在竞争者—侵占者模型中，私有财产均衡的存在和稳定也与这些因素紧密相关。

Hammerstein（1981）则发现了鹰—鸽模型的第三个问题，他认为参加竞争的行为者除了它们所采取的策略不同外还在其他方面有所区别。它们会在可能的特性例如力量、年龄、相对的格斗能力和需要等方面表现得不同。虽然在本文中我们只处理力量存在差别的情况，但是其他的情况同样也可以融合到一个竞争者—侵占者模型中。

第四个问题是鹰—鸽模型事实上预测了两个全局稳定中的任何一个都能将鹰—鸽的混合策略排除的情况。第一种策略是中庸策略，在这种策略下第一个占据领地的行为者将成为不需竞争的所有者。第二种策略被称为“共产主义的”策略，即当一个行为者在遇到一块已经被占据的领地时，在不发生争斗的情况下，就能赶走目前的占据者。由于第二种策略会导致无限侵占的情况出现，从而使得没有行为者能够从领地中得益，所以这种策略经常被否定。但是这种否定又是没有保证的，在竞争者—侵占者模型下，有很多组合理的参数值导致了共产主义者的而不是中庸的策略均衡出现。而我们也将会在第八部分给出一个合理的例子。

关于鹰—鸽模型的最后一个问题在于分析中所包括的策略选择是任意的。更为重要的是，竞争者的打斗能力被看作是一致的和外生的。相反地，考虑老鹰们把不同的资源用于提高打斗能力的情况

（Hammerstein，1981）。这时，如果在合理的条件下，只要存在一个正的概率让行为双方都错误地以为自己拥有这块领地，那么即使领地的目前所有者相对于潜在的侵犯者占有打斗上的优势，一个行为者扮演老鹰并且力量为 V 的中庸策略均衡也将被另一个行为者扮演老鹰并且力量为 $W < V$ 的中庸策略均衡所取代（参见第二部分）。

这些论据的关键不在于否认鹰—鸽模型的适用性，因为的确存在着其他的场合使得中庸策略不受这个问题的困扰。一种是错误对那些较弱的行为者而言成本更高，另一种则是更有威力的老鹰能以更大的概率和非常小的成本赢得这场战斗。在这些情况下，老鹰的力量会有增长的趋势，而这会抵消为了保持力量而付出的更多成本。但是无论怎样，竞争者—侵占者模型都已经拓展到足以包含这些情况的程度。

在竞争者—侵占者模型中，一个行为者必须作三个选择：是否在一块领地上投资，是否在遇到一块已经被占的领地时侵占该领地，以及当拥有领地时是否与怀有敌意的侵占者竞争。

这样在模型中将会存在 8 种纯策略：（a）被动行为者（Passive），即不投资也不侵占也不竞争；（b）投资者（Investor），进行投资但不侵占不竞争；（c）侵占者（Usurper），实施侵占但不投资也不竞争；（d）竞争者（Contester），不投资也不侵占，但参与竞争；（e）投资者—竞争者，参与投资和竞争，但不侵占；（f）投资者—侵占者，参与投资和侵占，但不竞争；（g）竞争者—侵占者，参与竞争和侵占，但不投资；（h）主动行为者（Aggressive），投资、侵占、竞争都参加。当我们不关心这 8 种策略的区别时，我们将把所有参与竞争的行为者称为竞争者，所有实施侵占的称为侵占者，所有进行投资的称为投资者。

在竞争者—侵占者模型中，竞争者和投资者—竞争者策略对应于鹰—鸽模型中的中庸策略，因为它们都不包括侵占行为但总是有竞争。而竞争者—侵占者策略和主动行为策略在格斗力量强时对应于老鹰策略，格斗力量弱时则对应鸽子策略，因为就像老鹰和鸽子一样，他们对所有者和非所有者都展开竞争。但是被动策略和投资者策略则

在鹰—鸽模型中没有相对应的策略，因为他们永远尊重其他行为者的私人财产，却不保护它们自己的财产。当所有的行为者不包括偶然的突变体采取竞争者或者投资者—竞争者策略，并且群体的数目也是演化稳定时，我们就可以说存在一个私有财产的均衡，因为绝大多数的行为者都尊重了所有权。[1]

是否选择进行投资对于模型的运行来说并不是严格必需的，但它可能出现在许多实证上很重要的例子中。在缺乏安全感的情况下，行为者没有动机通过投资来改善它们占据的领地，因为被其他行为者侵占的概率会非常高。私人财产在投资的回报比较高时才最可能得到推广，因为那些较小的，又相对封闭的尊重私人财产的群体会在牺牲不尊重私人财产的竞争者们利益的前提下获得使群体得以扩展的利益基础。道金斯（1982）可能是第一个认识到投资在一个动物环境中具有重要意义的科学家。这个主题在当代有关功能构造的文献中是非常突出的（Laland，Olding-Smee and Feldman，1999，2000；Laland and Feldman，2004）。事实上，正如我们所将看到的，相对于存在的可开发的领地而言，当群体密度较高时，私人财产才是进行高水平投资的前提条件。

在竞争者—侵占者模型的辅助下，我们指出在某些合理范围内的参数值下，私人财产均衡将会存在。但是私人财产的均衡却很少是全局稳定的，一部分竞争能力较强的竞争者将侵犯一个由各种类型的侵占者所组成的群体，而且一旦这种侵犯形成，他们将很难再被驱逐出去。

竞争者—侵占者模型的一个特点是竞争者和侵占者的力量由演化力量内生决定。一种力量是否是演化稳定的将取决于那些必须由模型来确定的生态和结构因素，它们包括自然和领地的动态变化、搜寻的形式以及由所有者和非所有者会面频率所决定的迁徙及各种策略所带来的成本与收益一览表。

在下文我们将要讨论的例子中，私人财产均衡背后的直观意义是非常简单的。在第一个例子中，因为有一个正的侵占成本 c_u 存在从而使

得在位者比侵入者更有优势，或者说一个所有者赢得竞争的概率 p_c 大于1/2。在第二个可能更具实证意义的例子中，无论是所有还是非所有者所能拥有的力量都是内在决定并取决于适应和选择的演化过程。在一个私人财产均衡中，行为者平均起来在拥有领土时比没有拥有领土时使用力量水平更高而且无论是所有者还是非所有者都没有动机改变这种力量水平。因此，一个所有者通常可以预期在竞争而不是放弃所有权时能获得收益，反之对非所有者也成立。这个均衡对应于博弈论中贝叶斯纳什均衡的概念（Gintis，2000），这是一个很少应用于动物行为模型但对理解私人财产非常关键的概念。

2　鹰—鸽博弈中的演化不稳定性

设想两个行为者遇到了一块价值为 $2r > 0$ 的领地。当双方都采取鸽子策略时，如果双方装模作样的（较小）成本为 $d < r$，那么各自的预期报酬即为 $r - d$，然后领地被随机地分配给两个竞争者中的任何一个。当一只鸽子和一只老鹰相遇时，鸽子得到零而老鹰得到 $2r$（整块领地）。老鹰们会付出 $2v > 2r$ 的力量来参加战斗。因而当两只老鹰相遇时，胜利者得到 $2r$ 而失败者失去 $2v$，所以每只老鹰平均得到 $r - v < 0$。检验这里的纳什均衡存在而且惟一是非常容易的，我们可以得到采取鸽子策略的可能 α^* 为：

$$\alpha^* = \frac{v - r}{v + d}$$

而每个行为者的收益为：

$$\pi^* = (r - d)\ \alpha^*$$

注意到当老鹰为竞争所付出的成本 v 越大时，均衡中采取老鹰策略的可能 $1 - \alpha$ 就越小，而行为者各自的收益就越大。

现在我们引入中庸策略，即当行为者第一个发现资源时就采取老鹰策略，当其第二个发现时则采用鸽子策略，同时每一事件发生的概率均为1/2。我们非常容易检验中庸策略可以侵入任何老鹰策略和鸽子策略所组成的群体，所以它是全局稳定的。

但是现在我们定义一只老鹰为 x，它将会把 $2x$ 的力量投入到斗争中。我们把标准鹰—鸽博弈中的老鹰标记为老鹰 v，然后我们增加一种老鹰 w 的策略，其中 $r < w < v$。当老鹰 w 和老鹰 v 相遇时，我们假定老鹰 w 获胜的可能性为 $f = w/(v + w)$，无论胜或负，每只老鹰花费比例为 μ 的自身力量在格斗上，除此以外当它失败时它还花费了比例为 $1-\mu$ 的力量。因此对老鹰 w 来说它的竞争成本为 $\mu w + 2(1-\mu)(1-f)w$，而对老鹰 v 的成本为 $\mu v + 2(1-\mu)fv$。[2] 现在我们定义一个记作 B_x 的中庸策略，当行为者先占据一块领地时就像老鹰 x 一样行为，否则就像鸽子一样行为，考虑一个由中庸策略 B_y 和 B_w 所组成的群体，$r < w < v$。因为任何中庸策略在面对其他策略时能取得 r 的收益，并且任何由中庸策略组成的混合策略本身是中性稳定的，因此没有中庸策略是演化稳定的。但是事实上，假如存在某个时间段，在其中行为要保持力量水平 $2v$ 的成本为 $c(v) > 0$，且 $c'(v) > 0$，或者两个行为者同时认为自己是一块领地的主人并且承认这个错误的成本随着老鹰力量的增长而增加，即使这种可能存在的概率非常小，也可以使中庸策略 B_v 严格劣于中庸策略 B_w。在这里我省略了显而易见的相关运算。我们发现即使当行为者竞争时，原居住者会比侵入者有更大的可能获胜的结论仍然成立。这是因为竞争中的每一个行为者成为原居住者或侵入者的可能性是相同的，所以如果他们都错误地扮演了原居住者，那么预期收益将是相等的。

3 田块与动物[3]

这里我们应用竞争者—侵占者模型的生态学背景为一块由许多可分

割的田块组成的土地，其中每个田块都是不可再分的，因此只能有一个所有者。当然土地所有者可以获得由该田块带来的全部收益。而非所有者当遇到一块田块时它们可能试图侵占该田块，也可能干脆继续寻找。在每一段时期内每一块田块都有一定的概率“死亡”（即变得不再肥沃），而且会在一段给定的时间内保持休耕状态，在此期间无法从该田块里得到任何收获。过了休耕阶段以后，注意到该田块会有一固定的概率在接下来的时间段内恢复生机。田块死亡和繁荣的周期足够长，以至于在一块死亡的田块上等待其复活是不值得的。

我们设想领地有 n_p 块不可分的田块组成，每块田块要么处于繁荣状态要么处于死亡状态。时间被分割为等长的几个阶段，田块在每一个阶段死亡的概率为 p，而死亡的田块在经历了 k_f 个阶段后，其后每个阶段恢复生机的可能为 q。在死亡复活前所历经的 k_f 段时间中，任何一段都被称为田块的休耕期。

我们作这些简化假设的目的在于得到一个具有高度操作性的分析模型。它很容易加上空间维度，季节，田块间的相关性，田块的可分割性和其他能反映领地重要属性的特征。

设想肥沃田块的一般密度为 f，已经休耕了 k 年，$k=1, \cdots, k_f$，田块的预期数量是 fp，所以预期中比例为 fpk_f 的田块在每个阶段将会休耕。假如有比例为 g 的死亡田块既不处于休耕阶段也没有恢复生机，那么一般地我们有 $f+fpk_f+g=1$。同时也有 $fp=qg$，因为重新恢复的概率 q 在稳定的肥沃率 f 相一致的前提下，必须满足上述的等式。所以我们能得到 q 的值：

$$q=\frac{fp}{1-f\ (1+pk_f)} \tag{1}$$

假设这里有 n_c 只生物，我们将把它们称为动物（critters），它们中的一些拥有肥沃的田块。在每个阶段结束时，那些死亡田块的所有者还有无产者将会在其他地方寻找生活资料的过程中被随机分配到另一田块。如果它们能找到无主的肥沃田块，那么它们就成为该田块的主

人。而假如它们找到的田块已被其他行为者拥有，那么新来者可以试图侵占、替换目前的所有者。如果一场竞争的失败者生存下来了，我们假定它离开了该田块而且不会与同一个对手再进行竞争。

4 动物的适存度

让 π_g 代表一只动物在一肥沃田块上为改进土地而进行投资（假如该动物进行投资的话）所获得的适存度值（整个生命中孕育的后代总数目）。类似地，让 π_b 代表无产动物的适存度。考虑在没有投资的情况下，b_o 是每一阶段在肥沃田块上的后代的繁殖率，而 c_m 时是迁徙到另一田块所付出的成本。在本文中，成本与行动相联系的死亡率是对等的，而收益是以后代数目来衡量的。假如一田块拥有者投资 c_i，那么它在每一阶段将享受一份数量为 b_i 的额外回报。在我们的计算中，我们假定投资能促进适存度的增加。当不符合这种情况时，等式 $c_i = b_i = 0$ 始终成立。我们得到：

$$\pi_g = b_o + b_i + \pi_g (1-p) + \pi_b p (1-c_m) \tag{2}$$

因为一个所有者再生产的概率为 $b_o + b_i$，有 $1-p$ 的概率保持土地的肥沃而且有 p 的概率遭受可能性为 c_m 的死亡，而假如这种可能性被避免，那么留下的处于无产状态的行为者的现值为 π_b。注意到当下一阶段田块继续保持肥沃时，所有者将会有与现期同样的期望收益，这是因为我们假定田块死亡的概率是固定不变的。去掉这个假定会使模型复杂化，但是能够使我们进一步研究田块的使用时间以及所有者和侵占者所掌握的不同知识将会怎样影响侵占和竞争的成本与收益。我们不会在本文中继续讨论模型这方面的内容。

我们也得到：

$$\pi_b = f_f (1-c_i) \pi_g + (1-f_f) (1-c_m) \pi_b \tag{3}$$

这里f_f是一个迁徙者发现一肥沃无主田块的几率。等式（3）成立是因为迁徙者在投资阶段生存下来的可能性为$1-c_i$，然后进入所有者状态，或者说该迁徙者有$1-f_f$的机会吸收迁徙成本c_m，如果成本c_m保留下来，那么迁徙者还继续处于非所有者的状态。

注意，这假定了处于非所有者状态的收益相对时间的变化而言是固定的。一个更为复杂的模型将把季节性影响和动物的年龄作为行为者交互活动的部分决定因素纳入讨论范围。在本文中我们将不会继续这些争论。

联立等式（2）和（3），我们得到π_g和π_b的均衡解：

$$\pi_g(c_i, b_i) = \frac{(c_m(1-f_f)+f_f)(b_o+b_i)}{p(f_f c_i + c_m(1-f_f c_i))} \tag{4}$$

$$\pi_b(c_i, b_i) = \frac{f_f(1-c_i)(b_o+b_i)}{p(f_f c_i + c_m(1-f_f c_i))} \tag{5}$$

如果$\pi_g(c_i, b_i)(1-c_i) \geqslant \pi_g(0, 0)$，那么投资就会比不投资有更高的收益。从这里，找到了能满足投资的最小回报量$r_{\min}$：

$$r_{\min} = \frac{b_o(c_m(1-f_f)+f_f)c_i}{c_m(1-c_i)} \tag{6}$$

为了完整估计拥有一块肥沃田块的收益值，我们必须确定f_f的值。让f_o和f_p分别代表所有者和有主的肥沃田块的比例，于是有：

$$n_p f f_p = n_c f_o$$

这一等式反映了有主田块的数量与所有者的数量相等的事实（一只动物最多只能拥有一块土地）。我们也可以得到：

$$n_c(1-f_o)f_f = n_p f f_p p$$

这表示了在均衡下，拥有肥沃土地但却面临土地死亡的所有者数量等于找到肥沃土地的非所有者的数量。我们令$\alpha = n_c/n_p$，代表了每块田块的动物数量，由此我们可以得到：

$$f_o = \frac{ff_p}{\alpha} \tag{7}$$

$$f_f = \frac{f_o p}{1 - f_o\ (1 - p)} \tag{8}$$

通过直接分析一块田块处于肥沃但无主状态的时间的预期比例，则 f_p 必须满足下面的等式 [4]：

$$f_p = \frac{1 - e^{-(\alpha - ff_p)}}{1 - (1 - p)\ e^{-(\alpha - ff_p)}} \tag{9}$$

在封闭形式下这个方程是无法解的，但是它所依赖的潜在参数的方向性比较容易得到，而且数字解也是非常直接的。

5 侵　占

假设有比例为 f_u 的动物在到达一块有主的肥沃田块时会采取强制性替换所有者的侵占策略。又假定原所有者不竞争，且侵占一个所有者的成本 c_u 足够的小，从而使得侵占的收益为正。因此，侵占者的存在会减小动物们投资于它们的领地的激励，即使在这样的投资可以提高任何拥有该领地动物的适存度的前提下也是如此。在这样的情况下，虽然投资和防止侵占“会对整个种群都有益处”，但却是不会发生的。

典型的私有财产无法维持的情况将会导致行为者放弃能增加收益的投资行为；而图 1 则说明了生态位构架（niche constraclion）不存在时的状况。这张图表示了通过计算机模拟的一片由 900 个田块组成田地，而它最初时的群体密度为一份地块上有一只动物。每一阶段有 20%的地块死亡，也有 20%的肥沃土地重新恢复到最初的状态。休耕期设为五个阶段。从（1）式我们可以得到当休耕期后田块恢复肥沃状态的概率为 6.67%时，土地的肥沃率将长期保持基本不变。

最初分配给未改善土地所有者的收益为10%（意味着该田块所有者在每个阶段有10%的机会可以拥有一个孩子），但是这个数据在上下每100个时间段后都要调整1个百分点，这是因为实际动物的数目会比均衡数900更大些或更小些。寻找的成本，即转移到另一块毗邻的田块的成本，定为0.03，在一块田块上投资的成本确定为0.20，而每个阶段给最初投资的回报率为60%。侵占的成本被定为0.08这一非常高的水平，以此来阻止侵占行为立即盛行的可能。同时也为了推迟侵占行为的出现，突变率被设定得非常低（每阶段0.000 05%，且只对新生的行为者适用），而最初的群体只是由被动行为者和投资者所组成。[5]

我们假定整个群体只进行单一的再生育，即在突变成其他行为者类型的可能性非常小的前提下，每个新生的行为者继承了父母的行为策略和其他相关特征。所有成本（即迁徙的成本，侵占的成本，投资的成本）都等同于一动物立即死亡的概率。因此，一旦所有者决定进行投资，且投资的成本为0.1，那么所有者有十分之一的可能立即死亡，否则它就可以一直享受它劳动的成果直到土地死亡。一块价值为 b_0 的田块允许它的所有者以 b_0 概率在每一阶段再生育。在图1中我们看到尽管为侵占策略设置了非常多的障碍，侵占策略还是出现了而且在几千个时间段后就很快地成为最为主要的行为策略。

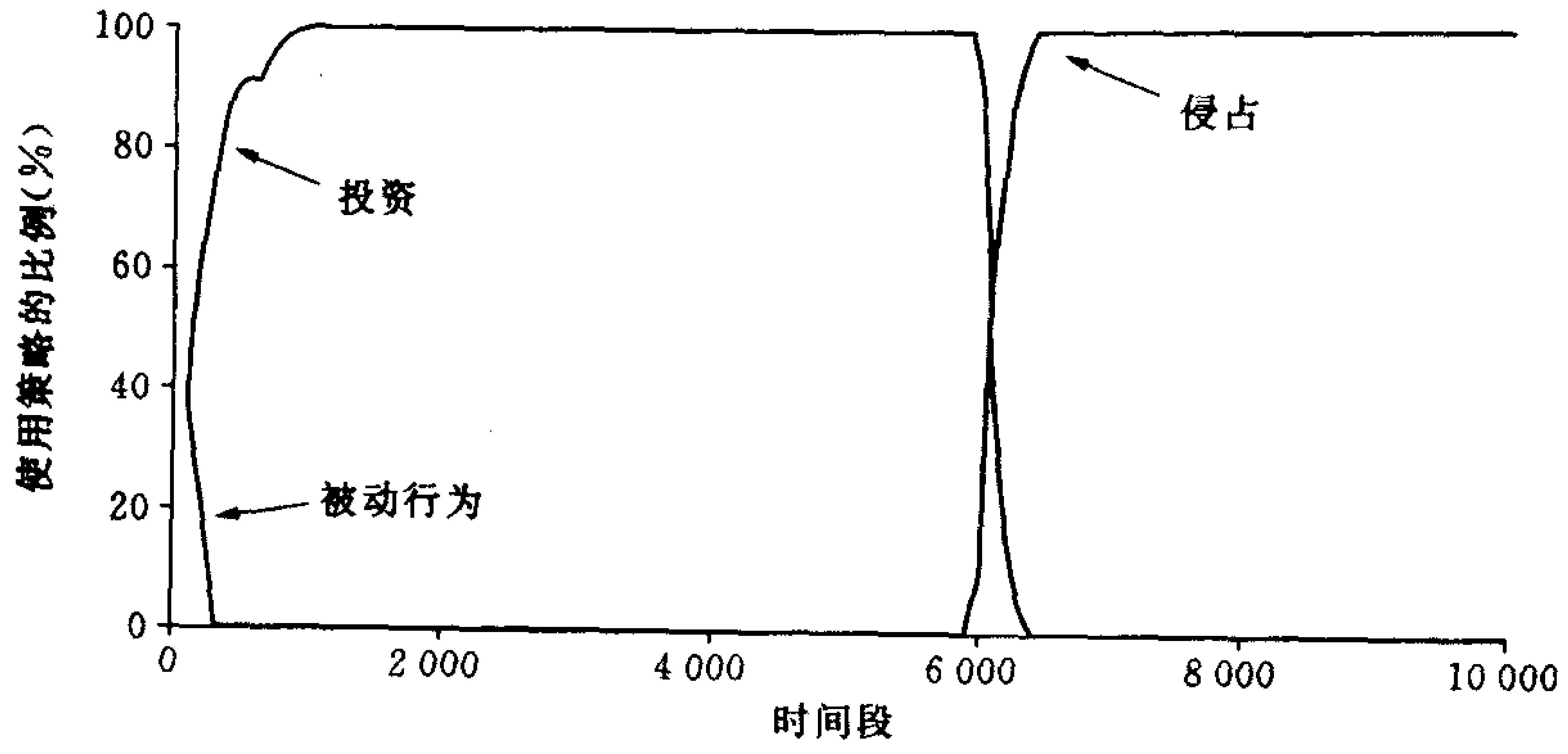

图1　侵占逐出投资

6 竞　　争

现在我们增加假设，当一个所有者被侵占时会跟对方展开竞争，从而在所有者和侵占者之间上演一场战斗。如果 c_c 为外在给定的竞争成本，而且对双方来说是相等的，那么竞争者中的一个就有 c_c 的概率死亡。如果双方都活了下来，则各有 50%的机会占有或保有财产。我们也假定在任何侵占发生之前，一切在田块上的投资都是所有者作出的，另外所有者的投资决定只依赖于它自身的策略，因而如果投资者是投资、侵占—投资、竞争—投资或主动型的行为者，那么它就会进行投资。特别地，投资决定并不依赖于在田块上非所有者的数量。虽然我并不觉得这一假设有什么不合理之处，但是我也没有正式探讨过这个问题。

对一个已经投资过的所有者来说，要使它参与竞争之后增加收益，就必须满足：

$$(1-c_c)\left(\frac{\pi_g+\pi_b}{2}\right)+\frac{c_c}{2}\pi_g>\pi_b$$

这是因为如果所有者不竞争，那么他得到 π_b（不等式右边部分）。而如果他参与竞争，那么侵占者有$\frac{c_c}{2}$的机会死亡且得到 π_g，或者有 $1-c_c$ 的可能双方都不死且他得到 π_g和 π_b的概率是同样的。此外，只有当下述条件满足时，一个非所有者才会通过不参加侵占增加适存度：

$$(1-c_u)\left((1-f_c)\pi_g+f_c\left(c_c\frac{\pi_g}{2}+(1-c_c)\left(\frac{\pi_g+\pi_b}{2}\right)\right)\right)<\pi_b$$

这里 f_c 是所有者 1 中参加竞争的比例。为了理解这一条件，请注意一个侵占者会立刻支付 c_u 的成本，而且如果他存活下来，他会因为目前的所有者不竞争而得到这块田块（这一事件发生的概率为（1 −

f_c））。否则的话，$\frac{c_c}{2}$的可能是因为所有者死亡而占有财产，$1-c_c$的可能双方都活下来并且侵占者获得田块的概率为1/2。下面将提供一个充要条件，使得存在一个竞争成本c_c，对所有者而言参与竞争是有利可图的，而对非所有者来说实施侵占却是无利可图的。因此，

$$\frac{\pi_g}{\pi_b} < \frac{1-(1-c_u)f_c}{(1-c_u)(1-f_c)} \tag{10}$$

当且仅当下述条件下将存在一个竞争成本c_c，使得尊重私有财产能增加收益：

$$f_c \geqslant f_{c\min} = \frac{\pi_g(1-c_u)-\pi_b}{(1-c_u)(\pi_g-\pi_b)} \tag{11}$$

从（11）得到一个使得私有财产均衡存在的必要条件是$c_u > 0$：即使在无竞争状态下，这里也会有一些不对称，使得侵占的成本很高。在很多情况下，这个条件是合理的，因为动物们即使不会参与一场高度减少收益的竞争，但仍可能会进行一场小规模战斗。在更多的情况下，这个条件是不合理的。这只是帮助说明了在动物群体中私有财产相对较为稀有的现象。

因为$f_{c\min} > 0$意味着$\pi_g > \pi_b$，从（11）得到的第二个结论是侵占者均衡总是演化稳定的，除非侵占本身并不能提高收益。因此就会有一部分严格大于零的竞争者侵犯这样一个群体，事实上，给各种参数配上的合理数值有可能使得这个比例很大。举例说，我们计算$c_u = 0.2$和$\pi_b/\pi_g = 0.5$的情况，得到$f_{c\min} = 75\%$。[6]

图2是一个演化稳定的私有财产均衡的计算机模拟。这个模型的参数与前面一样，只是我们把迁徙成本从0.03改成了0.02，而这会减少π_g和π_b之间的差距，然后我们将竞争成本增加到0.75，这意味着竞争者有37.5%的机会死在一场竞争中。在这种情况下，假如我们又允许被动和投资型策略，因此侵占或者竞争是不可能的，那么从任何最初的群体状态出发，一个投资者均衡都会出现。现在我们允许有侵占和

投资—侵占型策略，也是从任何最初的群体分配状况出发。侵占会降低投资的回报，所以，投资行为减少，而且最初只剩下侵占策略。当竞争进入模型时，从一高水平的投资—竞争型行为者群体出发，私有财产均衡将被永远保持。但是这一均衡并不是全局稳定的，因为如果我们一开始的时候竞争者太少，对应于鹰—鸽模型中的共产主义策略的侵占均衡将不能被排除，除非突变所带来的一系列非常不可能的结果大大增加了竞争者的比例，即使他们的适存度很低。

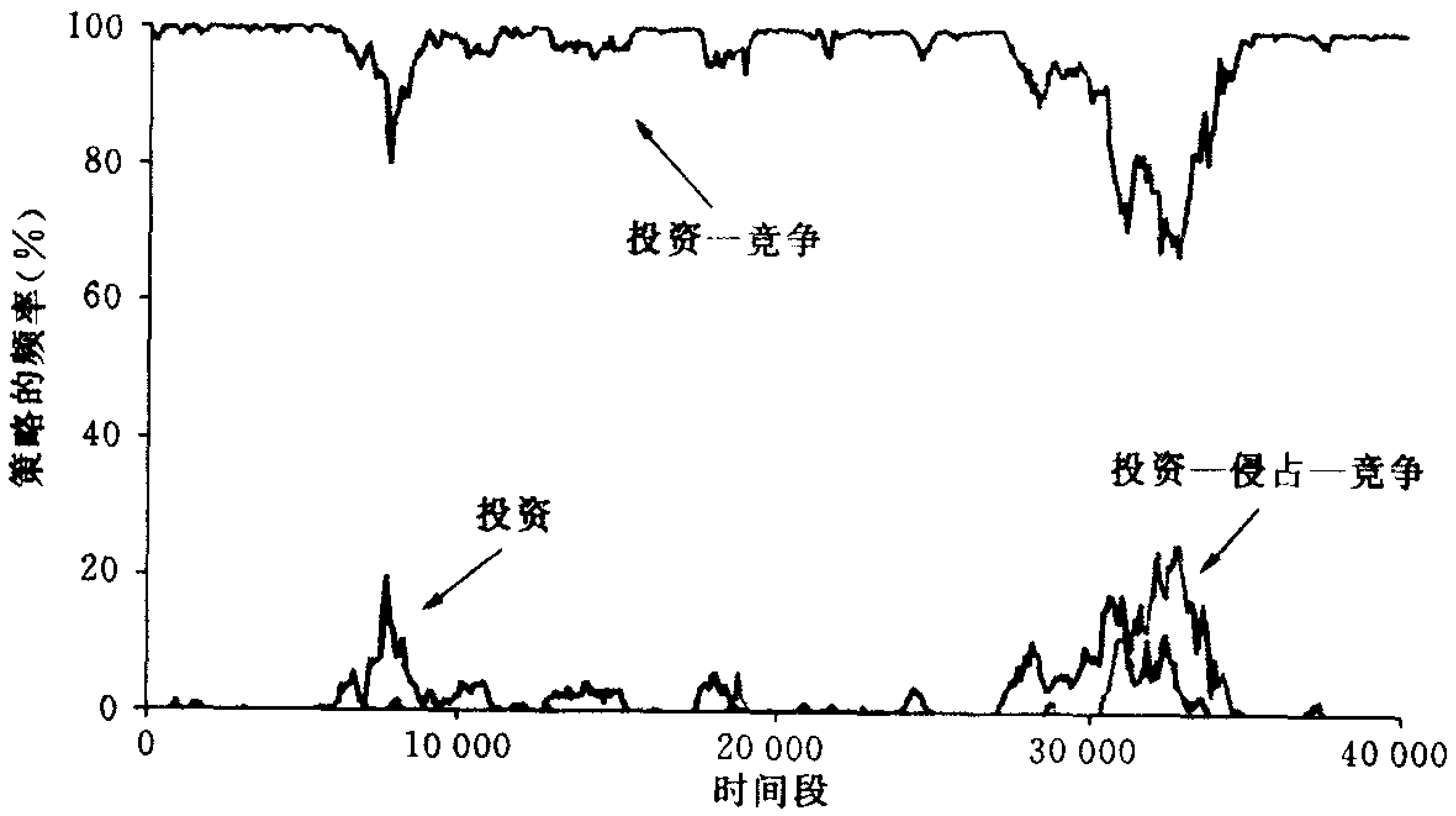

图 2　一个私人产权均衡

7　内生的战斗力

对（11）式中所含参数合理取值范围的估计使得我们认识到放在私有财产均衡上的注意力一般是相当有限的。正如我们已经看到的，在合理的条件下，为了维持一个私有财产均衡，群体中至少必须有75%成员是竞争者。这并没有反映私有财产的现状，一般私有财产出现得相对较少，但是一旦存在就会倾向于比较稳定的状态。此外，我们要求以形式为 $c_u > 0$ 出现的不对称或者私有财产不稳定的情况。除了在一

些特殊的例子中，这里并不存在合理的原因使得这种不对称成立。

从我们的模型中缺失的正是私有财产最为核心的性质：所有者配置了足够的格斗力，以至于它不会让潜在的侵占者加入到这场斗争中来。

假设每个动物最初都拥有一对承诺，一个说明当它是所有者时的斗争强度，而另一个说明当它侵占非所有者时的斗争强度。我们可以把这两者对应地称为防御力量和进攻力量。

设想当一个侵占者和一个竞争性的所有者相遇时，侵占者的进攻力量为 s_u，所有者的防御力量为 s_o，这里 $0 < s_u, s_o \leqslant 1$，然后让 $p_u = s_u/(s_u + s_o)$。我们假定在 $s = (s_o + s_u)/2$ 的概率下，这场会面会以斗争者双方中的一个死亡而结束，而所有者的死亡概率为 p_u，侵占者死亡概率为 $1 - p_u$。在这种情况下，幸存者保有或占有这片田块的所有权。假如双方都存活下来，则侵占者成功替代原所有者的概率为 p_u。注意到双方以投入战斗的力量来表示最终拥有领地的概率。这组跟竞争成本有关的假设中肯定不会存在什么东西是神圣不可侵犯的。在某些类型的战斗中，死亡的概率用最小的进攻和防御力量来表示更好，而且这将非常容易想像其非线性成本和收益表。

在让一个所有者参与竞争并且加上新假设的情况下，要满足的条件为：

$$s(1 - p_u)\pi_g + (1 - s)((1 - p_u)\pi_g + p_u \pi_b) > \pi_b$$

第一项是侵入者死亡而所有者存活的概率，乘以所有权的价值。而第二项是都没有死亡的概率，乘以在一场竞争后用目前所有者的预期收益来表示的值。现在侵占的条件为：

$$(1 - f_c)\pi_g + f_c(sp_u \pi_g + (1 - s)(p_u \pi_g + (1 - p_u)\pi_b)) > \pi_b$$

这里 f_c 是所有者参与竞争的可能性，这些等式可以化为：

$$\frac{\pi_g - \pi_b}{\pi_b} > \frac{s_u}{2}\left(1 + \frac{s_u}{s_o}\right) \tag{12}$$

对所有者而言有：

$$\frac{\pi_g - \pi_b}{\pi_b} > f_c \frac{s_o}{2} \frac{s_u + s_o}{(1-f)_c s_o + s_u} \tag{13}$$

对潜在的侵占者而言，当第一个不等式成立而第二个不成立时，一个私有财产的均衡就会出现。

不等式（12）和（13）说明了模型的两个重要性质。第一，私有财产不能是全局稳定的，这是因为当 f_c 很小时，第 2 个等式会成立。第二，在适当的条件下，所有者会有个人动机选择较大值的 s_o，而非所有者可能不会有动机相应地选择较大值的 s_u。为了理解这一条性质，我们注意到在两个等式中 s_o 和 s_u 的表达方式是有所区别的。在（12）中，所有者可以自由选择 s_o 的值，而右边部分则是在种群中实际进行侵占的非所有者的 s_u 值分布基础上得到的预期值。因此所有者可以选取某些 s_o 使得等式满足，只要 s_u 的均值不是太大。在（13）式中，潜在的侵占者可以选择 s_u 的值，但是右边部分是在所有竞争者的 s_o 值分布基础上得到的预期值。如果 f_c 足够小或者竞争者 s_o 的均值足够小的话，非所有者可以选择 s_u 的一个值使得侵占能够增加收益。但是如果 f_c 足够大且 s_o 也比较大，那么就存在使（13）成立的 s_u。特别地，如果 $f_c/(2-f_c) > (\pi_g - \pi_b)/\pi_b$，那么（13）式一定不成立。在这种情况下，潜在侵占者不会从选择较大的 s_u 中获益，而且即使 s_o 较小，第一个等式也会成立。这说明了在一个私有财产均衡中 s_o 和 s_u 是演化稳定的可能性。

但这种解释并不是正式的证明。事实上，要计算战斗力量 s_o 和 s_u 的均衡值，这个模型显得过于复杂了。然而以行为者基础模拟的这个模型在相当宽泛的初始条件下，证明了这些参数的演化是与一个私有财产均衡相容的。

在所有的模拟中，最初群体里的每一动物从单位间隔 [0.5, 1] 和 [0, 0.5] 的均匀分布中被相应地分配防御力和进攻力。新生的动物则从它们的父母那里继承这种力量，不过存在一个突变的概率，在此种情况下该力量大致会增加或减少 0.5 个百分点。因此这些打斗力量是完全

内生化的，除了强加到承受者身上的适存度成本，不存在任何东西能阻止侵占力量的快速增长，或者竞争者力量降低到较低的水平。正如我们所应该看到的，虽然在我们的模型中斗争力量是内生的，但是具有较大力量的竞争者可以在演化动态中存活下来，即使使得竞争者优于侵占者的不对称成本 c_u 被去掉时也是如此。虽然最终的私有财产均衡是非常稳定的而且正如在前面模型中所描述的那样，在使用合理的参数值上有很大的吸引力，但是这样的均衡并不是全局稳定的。

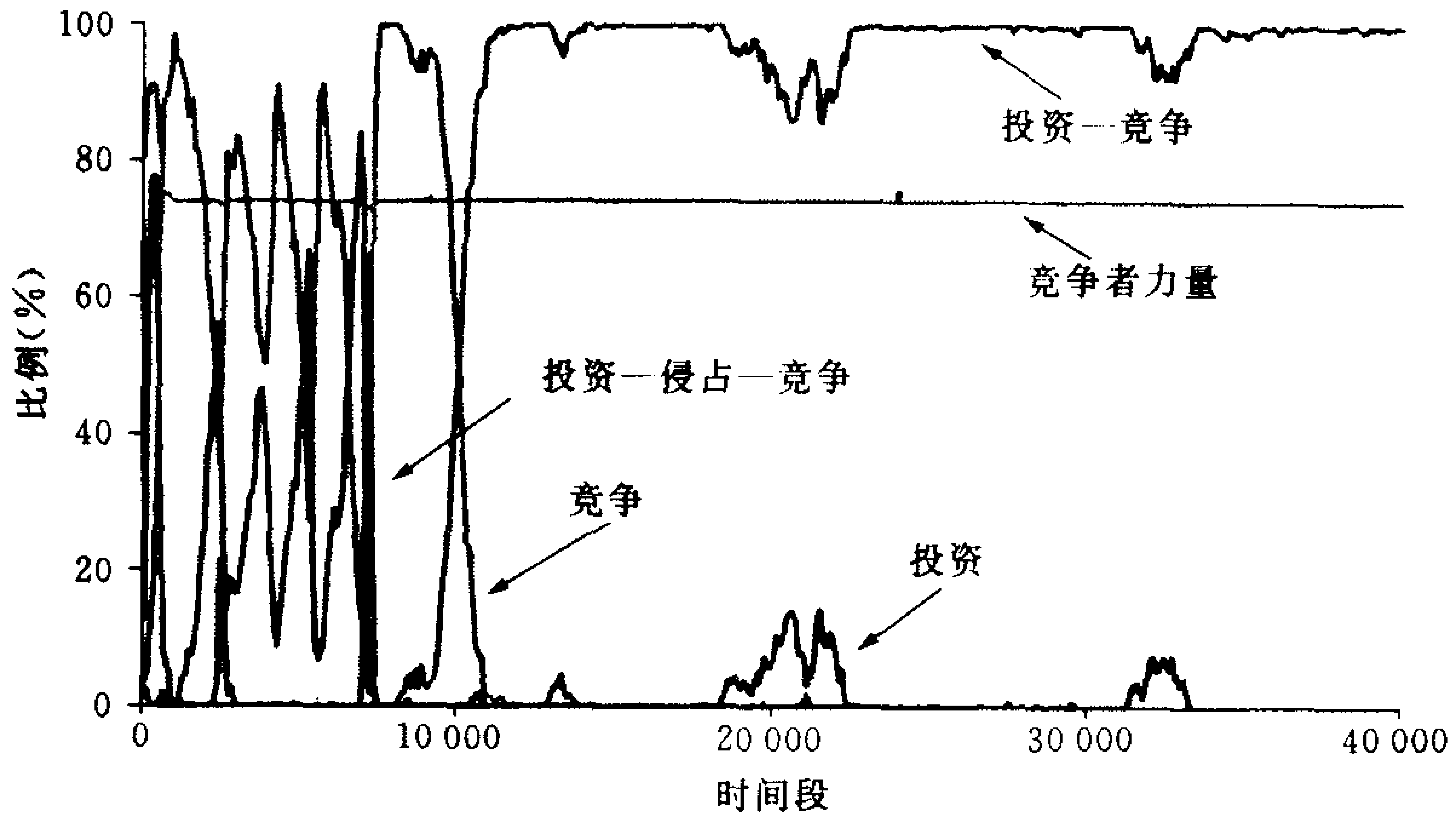

图 3　当竞争力量是内生时的中庸侵入

图 3 描述一个典型的模拟过程，其中最初投资—竞争的行为者比例非常低，只有 3.44%。投资—竞争型策略在一开始几乎就要消失了。但是，除了较小比例成为其他类型的突变体以外，最终会增长到一个固定值。对这一模拟的进一步研究得到在力量内生的情况下，可以构建出相当宽的模型参数值，使得私有财产均衡较为显著地在模型中出现。

8　一个反财产的均衡

考虑如下的一个情形：行为者都将死亡，除非它们能在每 n 天内至

少成功地拥有一块田块。一旦拥有田块以后，它们在每个时期能以 r 的速度进行再生产。当一个行为者遇到一块已经被占有的田块时，它对该田块的估价可能超过目前的所有者，这是因为一般来说侵入者比目前的所有者寻找另一块田块的时间更少，而且前所有者可以有整整 n 天的时间。在这种情况下，目前的所有者可能没有动机为保有田块而展开一场战斗，但是侵入者可能会。因此新来者可能在不进行战斗的情况下获得该田块。所以这里存在着一个稳定的侵占均衡，对应于鹰—鸽模型中的共产均衡。当然，在未来的一些时期，一个侵占者可能自身也会被侵占，但决不是说存在一个无效的所有权循环。

为了评价这样一个场景的合理性，请注意，假如 π_g 是一块田块所有者的适存度，然后 $\pi_b(k)$ 是一个还有 k 个时期可以去发现和开发田块的非所有者的适存度，于是我有了下列的递归方程：

$$\pi_b(0) = 0 \tag{14}$$

$$\pi_b(k) = f_f\pi_g + (1-f_f)\pi_b(k-1) \quad k = 1, \cdots, n \tag{15}$$

f_f 从一个非所有者变成一块田块所有者的概率，或者是因为田块不再被拥有，或者是非所有者可以用比较低的成本侵犯所有者。我们可以通过解（14）和（15）得到：

$$\pi_b(k) = \pi_g(1-(1-f_f)^k) \quad k = 1, \cdots, n \tag{16}$$

注意到当 k 和 f_f 越大时，一个非所有者的收益也越大。我们也可以得到方程（17）：

$$\pi_g = r + (1-p)\pi_g + p\pi_g(n) \tag{17}$$

这里 p 是田块死亡或者所有者在较低成本下被一个非所有者侵占的概率。我们通过解方程（17）得到：

$$\pi_g = \frac{r}{p(1-f_f)^n} \tag{18}$$

注意当 r 越大，p 越小，f_f 越大，n 越大时，一个所有者的收益越大。

正如在前面的模型中描述的，我们假定侵占者有 s_u 的进攻力量，竞争者有 s_o 的防御力量，这里 $0 < s_u$，$s_o \leqslant 1$ 且 $p_u = s_u/(s_u + s_o)$。在 $s = (s_o + s_u)/2$ 概率下一个行为者会死亡，竞争者以概率 p_u 死亡，侵占者死亡的概率则为 $(1 - p_u)$，假如双方都存活下来，侵占者成功替代所有者的概率为 p_u。我们假定有比例为 f_c 的所有者是竞争者，然后得到一个所有者和一个发现所有者田块的非所有者符合侵占策略的条件（即侵入者成功侵占，所有者不竞争）。当这些条件满足时，我们有 $f_c = 0$。

令 π_c为竞争而不是简单放弃田块所得到的适存度值（fitness value）。于是我们有：

$$\pi_c = s(1 - p_u)\pi_g + (1 - s)((1 - p_u)\pi_g + p_u\pi_b(n)) - \pi_b(n)$$

通过化简得到：

$$\pi_c = \frac{\pi_g}{2}\left(\frac{s_u^2 + s_o(2 + s_u)}{s_o + s_u}(1 - f_f)^n - s_u\right) \tag{19}$$

而且，π_c在s_o中递增，所以假如所有者竞争，他将确定 $\sigma_o = 1$，于是对所有者而言竞争能增加收益的条件变为：

$$\frac{s_u + 2/s_u + 1}{1 + s_u}(1 - f_f)^n > 1 \tag{20}$$

现在令 $\pi_u(k)$ 为一个必须在 k 段时期前拥有一块田块而且遇上一块有主肥沃田块的非所有者的适存度。那么行为者实施侵占收益为：

$$\pi_u(k) = (1 - f_c)\pi_g + f_c(sp_u\pi_g + (1 - s)(p_u\pi_g + (1 - p_u)\pi_b(k - 1))) - \pi_b(k - 1)$$

这个等式中的第一项是所有者不竞争的概率，乘以这种情况发生时侵入者的收益。第二项是所有者竞争的概率，乘以它进行竞争的收益，而

最后一项则是不侵占的收益。我们可以将等式简化为：

$$\pi_u(k) = \pi_g \frac{s_o(1-f_c)+s_u}{s_o+s_u} \tag{21}$$

在 $f_c>0$ 的前提下这个等式总是正的，而且随着 s_u 增加而增加，随着 s_o 增加而减少。因此侵入者总是会选择 $s_u=1$。另外，正如我们所预期的，在 $f_c=0$ 的情况下，侵入者实施侵占的概率为1，所以 $\pi_u(k)=\pi_g$。在任何情况下，不管 f_c 取什么值，侵入者总是会实施侵占。所以一个不竞争同时存在全局稳定的侵占均衡的条件（20）变为：

$$2(1-f_f)^n<1 \tag{22}$$

在 f_f 或者 n 足够大的情况下都将满足（22）式的情况。当（22）式不成立时，就会存在一个全局稳定的侵占—竞争均衡。

虽然 Maynard Smith（1982）描述了 *Oecibus cuites* 蜘蛛的例子，其中的侵占者几乎总是不用战斗就能替代原来所有者，但是侵占（即共产）均衡在文献中并不是经常见到的。而在人类中，这种情况已经被称为“可容忍的盗贼”，而且在若干狩猎—采集的群体中也能较频繁地被观察到（Blurton Jones，1987；Hawkes，1993；Bliege Bird and Bird，1997；wilson，1998）。更为非正式地，我观察了每个夏天我的鸟在进食和洗浴过程中的例子。一只鸟到来以后，会进食或洗浴一会儿，然后无抗议地被另一只鸟所替代，接着一切继续。这看起来像是在进食或洗浴了一会儿以后，就不再值得花任何精力来保卫这块领地。而侵占—竞争均衡可能用霍布斯均衡来命名更为恰当，因为这里有一场所有行为者针对所有行为者的战争，而且生命是野蛮和短暂的。（注意到这个例子中的情况，每一个对手都有 1/2 的概率会死亡。）

9 结　论

竞争者—侵占者模型预测私有财产均衡会很少出现，但是一旦出现

就会很稳定。当这样一个均衡确实存在时，它一般将依赖于所有者与侵入者之间的不对称，但是这种不对称可能完全是全局的统计意义上的，因为当所有者平均地更愿意比侵入者投入更多力量时，这种区别就是演化稳定的。假如我们只是从局部互动来看不对称的根源，那么要想对这种类型有所理解是不太可能的。

自从 John Maynard Smith 在 25 年前引入博弈理论模型后，动物行为学家对此持有两种截然相反的态度。一方面这样的模型确实对整个学科起到了革命性的作用，提供了有序和具有分析力的支持。另一方面，这样的模型经常只比“玩具”的意义多一点，把握住的只是在这个领域观察到的丰富动态内容的一小部分而已。

像竞争者—侵占者模型这样的方法可能提供一种非常有用的调和途径，在保持分析的清晰性和易操作性的同时，在生态细节方面非常丰富，作出了关于领地的生命周期、迁徙和所有者与非所有者的互动形态等方面的具体假设。这样的模型对说明系统所需要的参数数量进行了一定的控制。而且，每一个这样的参数，包括田块死亡的概率、迁徙的成本、所有者和非所有者在肥沃与死亡的土地上的适存度，在理论上同时也可能在实践上有实证估计的能力。事实上，移动和互动的结构可以在改变后适应不同的现实环境。

注释：

[1] 一个采取被动策略或投资者策略的群体也尊重私有财产，但会被侵占者侵犯，所以并不是演化稳定的。

[2] 这些可能性和支付在 $v = w$ 时是减少鹰—鹰支付和可能性的惟一线性函数，并保证在 $w = 0$ 时老鹰 v 取胜。这个模型的分析不依赖于 μ 的值。

[3] critter 有小动物的意思，也有家畜的含义，用于贬义的话，也可称呼人，现暂译为动物。——译者注

[4] 这里和下文我们都会运用指数近似 $(1-1/n)^n = exp - n$，特别地当 n 非常小时这个结果相当精确。

[5] 所有在这篇文章中出现的模拟都是作者用 PASCAL 编程语言写成的，程序的正确性经过了检验，如果需要的话可以用各种理论上的偏离情况来检验模拟结果。除非另有说明，所有的模拟采用相同的参数。

[6] 一个相似的结论可以从假定所有者和非所有者之间的不对称使所有者更有可能赢得竞争。假设在一场竞争中，所有者死亡的概率为 $c_c(1-p_c)$ 而侵占者的死亡概率为 $c_c p_c$。如果双方都存活下来，所有者拥有田块的概率为 p_c，而侵占者拥有田块的概率

为 $1-p_c$。因而 $p_c>1/2$ 时反映了所有者占有优势。所以我们非常容易得到致使一个私有财产均衡需要的竞争者的最小比例为 $f_{cmin}=(1-p_c)/p_c$。那么特别当存在非常强的所有者偏见时，f_{cmin}的值也会相应地变得比较大。例如当 $p_c=2/3$ 时，则 $f_{cmin}=50\%$。

参考文献:

Bliege Bird, Rebecca L. and Douglas W. Bird, "Delayed Reciprocity and Tolerated Theft," *Current Anthropology* 38 (1997) :49—78.

Blurton Jones, Nicholas G., "Tolerated Theft: Suggestions about the Ecology and Evolution of Sharing, Hoarding, and Scrounging," *Social Science Information* 26, 1 (1987) :31—54.

Davies, N. B., "Territorial Defence in the Speckled Wood Butterfly (*Pararge Aegeria*): The Resident Always Wins," *Animal Behaviour* 26 (1978) :138—147.

Dawkins, Richard, *The Selfish Gene* (Oxford: Oxford University Press, 1976).

—, *The Extended Phenotype: The Gene as the Unit of Selection* (Oxford: Freeman, 1982).

— and H. J. Brockmann, "Do digger wasps commit the Concorde fallacy?", *Animal Behaviour* 28 (1980) :892—896.

Eason, P. K., G. A. Cobbs and K. G. Trinca, "The Use of Landmarks to Define Territorial Boundaries," *Animal Behaviour* 58 (1999) :85—91.

Ellis, Lee, "On the Rudiments of Possessions and Property," *Social Science Information* 24, 1 (1985) :113—143.

Furby, Lita, "The Origins and Early Development of Possessive Behavior," *Political Psychology* 2, 1 (1980) :30—42.

Gintis, Herbert, *Game Theory Evolving* (Princeton, NJ: Princeton University Press, 2000).

Grafen, Alan, "The Logic of Divisively Asymmetric Contests: Respect for Ownership and the Desperado Effect," *Animal Behavior* 35 (1987) :462—467.

Hammerstein, Peter, "The Role of Asymmetries in Animal Contests," *Animal Behaviour* 29 (1981) :193—205.

Hawkes, Kristen, "Why Hunter-Gatherers Work: An Ancient Version of the Problem of Public Goods," *Current Anthropology* 34, 4 (1993) :341—361.

Laland, Kevin and Marcus Feldman, *Niche Construction* (Princeton, NJ: Princeton University Press, 2004).

—, F. J. Olding-Smee and Marcus Feldman, "Evolutionary Consequences of Niche Construction and Their Implications for Ecology," *Proceedings of the National Academy of Sciences* 96 (1999) :10242—10247.

—, —, and —, "Group Selection: A Niche Construction Perspective," *Journal of Consciousness Studies* 7, 1/2 (2000) :221—224.

Maynard Smith, John, "Group Selection," *Quarterly Review of Biology* 51 (1976) : 277—283.

—, *Evolution and the Theory of Games* (Cambridge, UK: Cambridge University Press, 1982).

— and G. R. Price, "The Logic of Animal Conflict," *Nature* 246 (2 November 1973) :15—18.

Mesterton-Gibbons, Michael and Eldridge S. Adams, "Landmarks in Territory Partitioning," *The American Naturalist* 161, 5 (May 2003) :685—697.

Riechert, S. E., "Games Spiders Play: Behavioural Variability in Territorial Disputes," *Journal of Theoretical Biology* 84 (1978) :93—101.

Schlatter, Richard Bulger, *Private Property: History of an Idea* (New York: Russell & Russell, 1973).

Senar, J. C., M. Camerino, and N. B. Metcalfe, "Agonistic Interactions in Siskin Flocks: Why are Dominants Sometimes Subordinate?," *Behavioral Ecology and Sociobiology* 25 (1989) :141—145.

Stevens，Eliszbeth Franke，“Contests Between Bands of Feral Horses for Access to Fresh Water：The Resident Wins，” *Animal Behaviour* 36，6 (1988) ：1851—1853.

Williams，G.C.，*Adaptation and Natural Selection：A Critique of Some Current Evolutionary Thought* (Princeton，NJ：Princeton University Press，1966) .

Wilson，David Sloan，“Hunting，Sharing，and Multilevel Selection：the Tolerated Theft Model Revisited，” *Current Anthropology* 39 (1998) ：73—97.

不平等的遗传*

萨缪·鲍尔斯　赫伯特·金迪斯

关于政府在减少经济不平等方面应起何作用，人们的看法大相径庭。自利和价值观念的差异部分解释了人们在再分配问题上的争议。到目前为止，最困扰人们的问题是为什么穷者愈穷、富者愈富。统计数据表明无论富人和穷人都持有类似的想法，即那些认为“在人生中获得成功和领先他人”取决于“勤奋”或者“愿意承担风险”的人，他们往往反对再分配的方案。相反，那些认为成功的关键在于“从家庭继承财产”、“父母和家庭氛围”、“人际关系和认识恰当的人”或者“成为白人”的人则支持再分配方案（Fong，2001；Bowles，Fong and Gintis，2002a）。如果个人是由于继承而获得成功，那么即使是微小的成功，人们也会感到不公平。相反，通过别的方式获得成功的人即使取得巨大成就也不会招致别人的反感，只要竞争是公平的。

公平竞争怎样在代际间发挥作用？强调代际间经济地位转移的生

* 原文题目为 The Inheritance of Inequality，发表于 *Journal of Economic Perspective* 16，3（2002），pp.3—20，吴灵译。我们在这里要感谢 Jere Behrman，Anders Bjorklund，Kerwin Kofi Charles，Bradford DeLong，Williams Dickens，Marcus Feldman，James Heckman，Tom Hertz，Erik Hurst，Arjun Jayadev，Christopher Jencks，Alan Krueger，John Loehlin，Casey Mulligan，Suresh Naidu，Robert Plomin，Cecelia Rouse，Michael Waldman 和 Elisabeth Wood，感谢他们为本文所作的贡献。我们要感谢研究助理 Bridget Longridge 和 Bae Smith 所做的工作，感谢 John D. 和 Catherine T. MacArthur 基金会提供的资金支持。

成机制是什么？ 这些机制服从公共政策是不是在某种程度上使经济成功的取得更公平？ 这些是我们在本文中将尽力回答的问题。[1]

毫无疑问，一个出生在富裕家庭的孩子一般能够接受更多和更好的学校教育，并能从物质、文化、基因的遗传中受益。 但是直到最近，经济学家们一致认为在美国，成功是指每一代人中的胜利和失败。 从 Blau 和 Duncan （1967）开始，统计学上的早期研究发现父母与子女长大后的经济地位之间存在很弱的相关性，这也表明美国确实是一块“充满机会的土地”。 例如，美国学者 Becker 和 Tomes （1986）的研究表明，父母与子女的收入或者所得（或是它们的对数）之间的简单相关系数平均为 0.15，他们因此得出结论：“除了受歧视所害的家庭外，几乎所有从祖上继承的所得优势或所得劣势都将在三代内消除殆尽。” Becker （1988）在美国经济学协会的就职演讲上表达了这样一个被广泛赞同的观点：“高收入或者低收入从父辈转移到子辈的程度不高。” （p.10）

但是越来越多的研究表明，高水平的代际流动性的估计在度量时容易犯两类错误：在报告收入时会出错，特别是在要求个人回忆他们父母收入水平的时候容易出错；认为当前收入的暂时组合与永久性收入不相关。（Bowles， 1972； Bowles and Nelson， 1974； Atkinson，Maynard and Trinder， 1983； Solon， 1992； Zimmerman， 1992； Björklund，Jäntti，and Solon，即将发表）两代人收入间的高噪音信号比（noise-to-signal-ratio）降低了代际间的相关。 当这些错误被纠正后，代际间经济地位的相关性变得非常显著，它们的值是 Becker 和 Tomes （1986）的研究结果的 3 倍。

经济成功的代际转移估计值的高度一致激发了实证研究。 大多数研究者都同意的事实包括：与那些在同一个种族里随机挑选出的有相似年龄差的人之间的收入水平相比，兄弟之间的收入水平更为相似；同卵双胞胎与异卵双胞胎或非双胞胎相比，其收入水平更为相似；出生在富裕家庭的孩子能够接受更多和更好的学校教育；财富继承对子孙后代的富裕起很大作用。 在以上几点和其他实证规律的基础上，似乎可以得出这样的结论，即不同机制的组合能够解释经济地位的代际间转移，这包括雇员需要的认知

技能和非认知个性特征的遗传传递与文化转移、财富和高收入的团体成员资格如种族，以及地位较高家庭孩子享有的优秀教育和健康体魄。

然而，经济成功的代际转移依然是一个黑箱。我们发现尽管富裕父母较好的认知表现和教育程度非常重要，但最多只解释了经济地位代际间转移的一半原因。此外，当增加收入特征的遗传传递看上去起作用时，智商的遗传传递相对来说不太重要。

可能有人认为黑箱问题的存在是因为相对于测度父母与子女的收入或所得的方法而言，对干扰变量的测度方法较差。但是事实上并不是这么回事。对受教育年限和其他从学校里所获知识的测度方法，跟测度认知表现（cognitive performance）一样只存在很小的误差。采用更好的测度办法当然是有帮助的，但是我们测量智商的办法不太可能有很大的改进，而且最近用改进后的办法对学校质量所进行的测度，并没有给我们研究黑箱问题带来多大的启示。问题的根本并不在于我们测度正确变量时方法太差，而是在于我们完全忽略了其他一些重要的变量。这些变量究竟是什么？

许多经济学模型把个人收入当成是个人带到市场上的生产要素（例如认知功能和教育）的总回报。但事实上，任何影响收入以及使父母一子女间存在很强相似性的个人特征都会影响到代际间经济成功的转移。这些特征包括种族、地理位置、高度、外表、或其他生理特征、健康水平，以及个性。因此，与标准的研究方法相反，我们给予这些收入产生特征（income-generating characteristics）以适当的考虑，而这些特征通常不被认为是生产要素。在有关代际间经济地位转移的研究中，我们认为在认知技能和教育方面已经研究过头了，然而在财富、种族和非认知行为特征方面的研究还有待深入。

1 代际间经济地位转移的测度方法

经济地位可以用离散的类别来界定，例如社会等级，也可以用连

续变量来测量，如收入、所得或者财富。离散的方法允许一个丰富但很难总结的表述，使用相关社会等级中的过渡率来描绘代际间地位保持过程（Erikson and Goldthorpe，1992）。相比之下，连续的方法能够基于两代之间经济地位的相关关系给出一个简单的度量。此外，这种相关关系还能够分解，用以说明父母孩子经济地位相似性的各种生成机制。两种方法都是很有洞察力的，但是为了简单起见，我们主要采用连续的测度方法。考虑到数据的可获得性，我们用收入或所得作为计量单位，虽然对大多数应用来说收入（是更具包含性的办法）是更好的选择。

我们用下标 p 来表示父母，而 y 是指个人的经济地位，经过调整后它的平均值为 $\bar{y}$，y 在代际间是一个常数项，β_y也是一个常数项，ϵ_y 是一个与 y_p 无关的干扰项，因此

$$y - \bar{y} = \beta_y(y_p - \bar{y}) + \epsilon_y$$

后代经济地位的均差等于父母经济地位与均值之间的差乘以 β_y，再加上干扰项。相关系数 β_y是用来测度代际间的收入弹性。在后面的实际工作中，除非特别提到，收入、所得、财富和其他用来测度经济成功的变量，我们都是取它们的自然对数值进行衡量。因此，β_y表示父母的经济地位变动 1%时，子女的经济地位变动的百分比。我们用 $1-\beta_y$ 来测量经济地位均值对后代经济地位的影响，$1-\beta_y$ 叫做对均值的回归（regression to the mean），它表明了个人希望他们自己的经济地位比他们的父母更接近均值的意愿程度（Goldberger，1989）。

代际间收入弹性与代际间相关系数的关系可以由下式表明：

$$\rho_y = \beta_y \frac{\sigma_{y_p}}{\sigma_y}$$

σ_y是 y 的标准差。如果 y 是财富、收入或者所得的自然对数，那么它的标准差就是测量不平等的无计量单位（unit-free）方法。因此，如果不平等在代际间不变化，则 $\sigma_{y_p} = \sigma_y$，那么 $\rho_y = \beta_y$。然而，当

收入不平等提高的时候代际收入弹性超过了 ρ_y，而收入不平等降低时代际收入弹性比 ρ_y 要小。在效果上，代际间相关系数 ρ 会受收入分配的影响，而代际收入弹性却不会。另外，ρ^2 测度了这一代经济成功与上一代经济成功之间线性关系的差异。

Mulligan（1997）和 Solon（2000）已经给出了代际间收入弹性的估计办法。Mulligan 给出的估计均值分别为：消费 0.68；财富 0.50；收入 0.43；所得（或工资）0.34；学校教育年限 0.29。证据表明，收入在代际间的持续程度有混合的趋势。许多研究表明代际间经济地位的延续性会随着年龄的提高而提高，儿子的延续性要比女儿强，当多年的收入和所得加以平均时延续性也会变强。最近 Mazumder（即将出版）的一个研究对多年平均的重要性作了描述。他用美国社会保障管理局提供的数据来估计代际收入弹性，当他把子辈的所得加以 3 年平均，把父辈的所得加以两年平均，得出的弹性是 0.27；把父辈的所得加以 6 年平均时弹性增加到 0.47；而把父辈的所得加以 15 年平均时这一数字增加到 0.65。

对于许多出身寒门的孩子来说，这么大的代际弹性意味着他们要变得富裕的想法只不过是种幻想吗？代际相关弹性是一种平均测度，凭此我们难以知道出生于富裕家庭、中等家庭和贫寒家庭的孩子在经济上成功的概率。通过计算这种条件概率以及观察整个转移矩阵，我们可以有一个全面的了解。Hertz（2002）的研究结果可以用图表 1 加以说明，图中把父辈的收入在坐标上分成十等分（从左到右依次对应着从贫困到富裕），把成年子女的收入在另一坐标上也用同样的方法来排序。高度表示从父辈坐标转移到子辈坐标的可能性。

尽管 Hertz 在数据中用的收入代际相关系数为 0.42，但是这也表明出身贫困的孩子和出身富裕的孩子之间的生命轨迹有很明显的差异。图中的“双峰”体现了贫困与富裕（虽然我们不期望“富裕陷阱”这个术语会变得流行）之间的这种差异。在坐标上 D 点表示出生在最富裕家庭（十等分中的前十分之一）的孩子，长大后将有 22.9%的机会步入

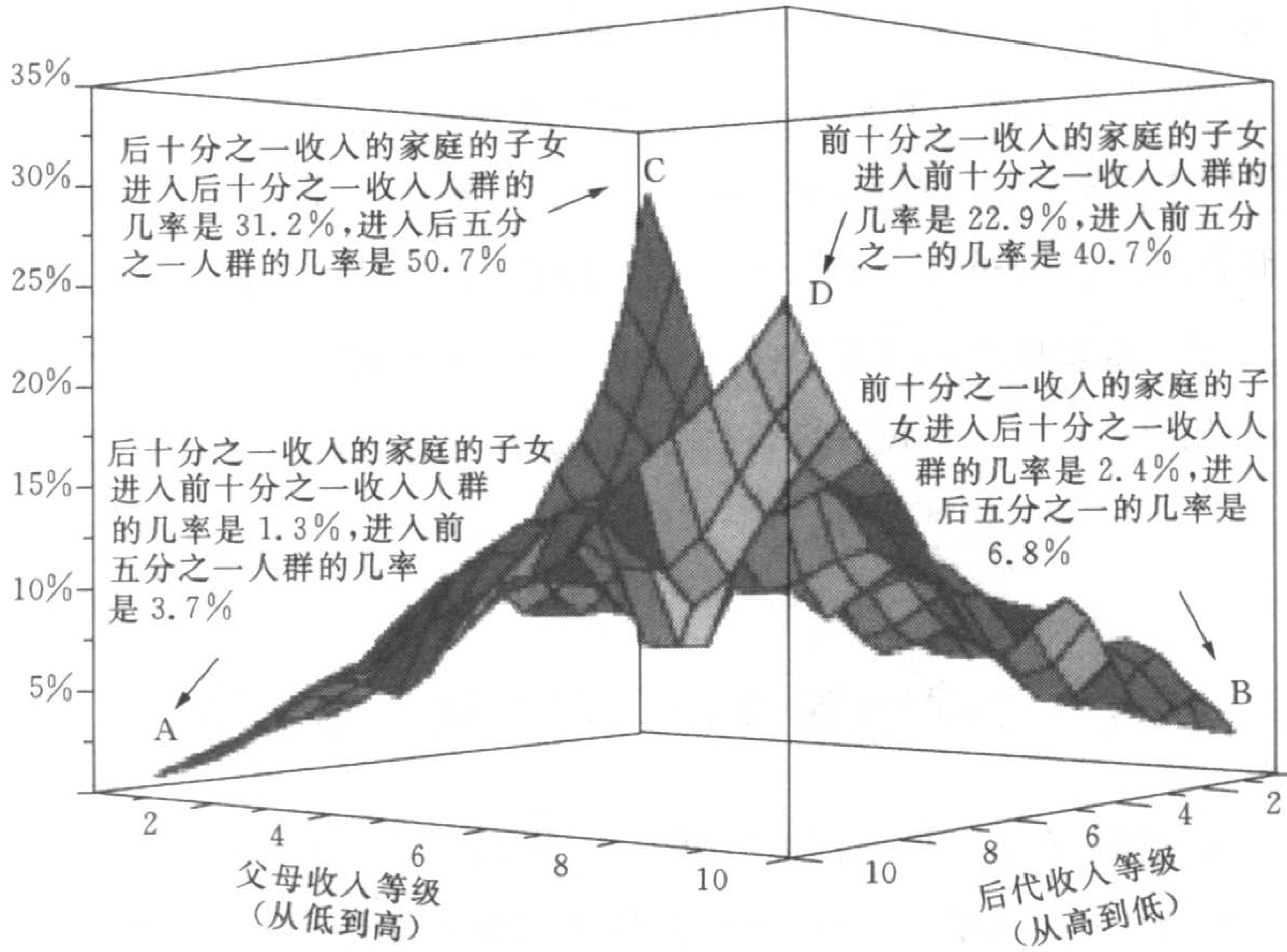

图 1　代际收入转移率

图中，细胞（i，j）的高度表示其父母亲的收入位于第 i 位置的人成年后将拥有位于第 j 位置的家庭收入的概率。在子女 26 岁以上开始衡量他们的收入，并且在观察到的数据里面把收入加以平均。在收入数据中，我们把收入年龄从 1 到 21 排列，得到平均数是 9.9。父母的收入是指当他们的子女在身边的时候的收入。我们把父母的收入也加以平均。在这些数据中，我们把收入年龄从 1 到 18 排列，得到平均数是 9.4。在图表中，经过简单年龄调整的父母与子女之间的收入相关系数为 0.42。

最富裕阶层，有 40.7%的机会步入比较富裕的阶层（十等分中的前五分之一）。A 点表示出生在最贫困家庭（十等分中的后十分之一）的孩子，长大后将有 1.3%的机会步入最富裕阶层，有 3.7%的机会步入比较富裕的阶层。C 点表示出生在最贫困家庭的孩子，长大后将有 31.2%的机会进入最贫困阶层，有 50.7%的机会进入比较贫困的阶层（十等分中的后五分之一）。B 点表示出生在最富裕家庭的孩子，长大后将有 2.4%的机会沦为最贫困阶层，有 6.8%的机会落入比较贫困的阶层。Hertz 的转移矩阵和其他的研究（Corak and Heisz，1999；Cooper，Durlauf and Johnson，1994；Hertz，2001）表明，在不同的

收入分配点上有不同的转移机制在发挥作用。例如财富遗赠在顶层的收入分配中可能起了很大的作用，而在底层的收入分配中，高犯罪率和身体健康因素可能又起到很大的作用。不同种族之间的动机模式也存在很大的区别（Hertz，2002）。从顶层沦落到底层的黑人是白人的五倍。因此成功的黑人并不像白人那样，能够有效地把成功转移给下一代。相应地，黑人从底层踏入上层社会的几率是白人的一半。

2 延续的源泉：文化，遗传和遗赠

经济地位在代际间有效地转移着。我们将试着去发现父母收入是如何影响子女收入的。我们把代际相关（或者说是代际收入弹性）分解成反映生成机制贡献的各个部分。这样我们就可以得出一些结论，例如经济地位代际相关的原因有一部分可以用智商遗传来解释，或者孩子长大后富裕是因为生于富裕家庭。

相关系数可以做到这一点。此外，我们用到的方法并不要求以特殊的次序来引进变量。假定父母的收入（用它的对数值来测度，用 y_p 表示）和子女的教育（s）影响到子女的收入（同样用它的对数值来测度，用 y 来表示）。跟其他相关系数一样，代际相关系数 $r_{y_p y}$ 测度了在多元回归预测 y 中，关于父母收入（$\beta_{y_p y}$）和子女教育（β_{ys}）的回归系数的总和。父母收入（$\beta_{y_p y}$）和子女教育（β_{ys}）都跟 y_p 与回归因子（当然，对于父母收入这个变量来说自相关为 1）相乘。正态回归系数只是在因变量、标准单位离差中发生变化，并与自变量发生的一个标准离差变化有联系。在这个回归分析中，父母收入的正态回归系数体现的是**直接作用**（direct effect）。在关于代际相关的分解式中，教育因素被称为**间接作用**（indirect effect）。图 2 [2] 说明了这样的分类：

$$r_{y_p y} = \beta_{y_p y} + r_{y_p s}\beta_{ys}$$

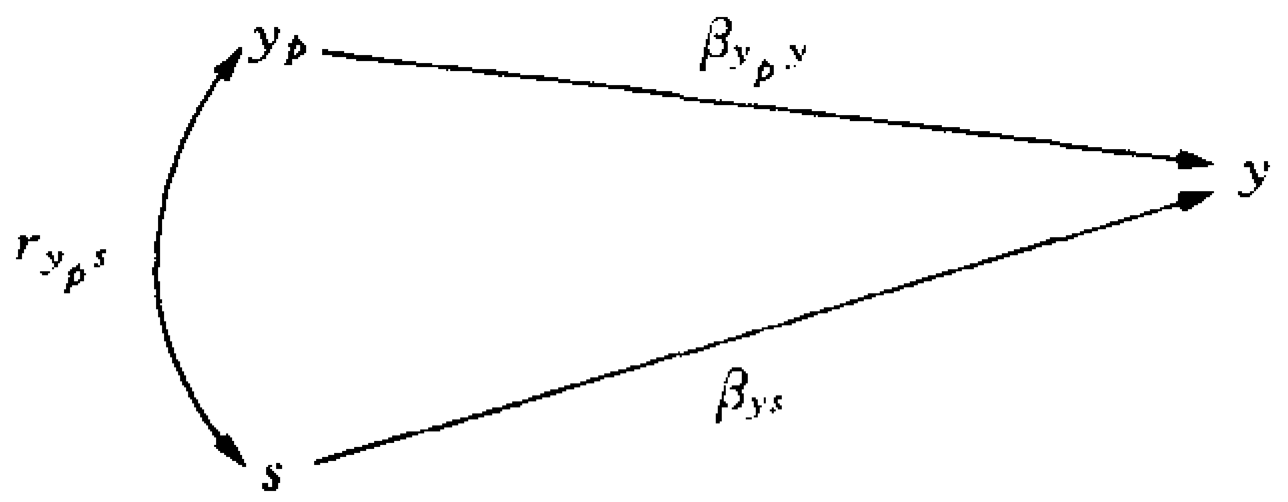

图 2　直接作用和间接作用总和的相关性

只要多元回归系数是无偏的，那么分解式中无论哪种变量间的相关系数都是有效的。特别，这里并不要求回归因子是不相关的。这个分解式让我们能够更加精确地了解引言里所提到的“黑箱”。当我们说标准的学校教育、认知水平和其他一些变量对观察到的父母子女间收入相似性的影响小于一半时，我们的意思是说在使用这种比较方法的许多研究中父母的直接作用不到代际相关性的一半（Bowles，1972；Bowles and Nelson，1974；Atkinson et al.，1983；Mulligan，1997）。

我们的策略是估计这些直接影响和间接影响的大小。分解式中，父母收入和其他变量（例如上面所说的学校教育）的相关关系被认为是与收入差异产生过程相联系的。当然这些相关关系并不需要反映因果关系。但是以上所提到的分解式能够重复地被用来解释父母收入和后代收入之间的相关关系，有时候是因果关系。例如，在研究财富转移过程中财富的作用时我们会问到为什么父母收入与子女财富是相关的。是因为遗产和身后物的转移，抑或是由于储蓄行为的文化传递造就了这种相关性？难道仅仅因为不知道父母和子孙的财富存在着什么样的相关关系，我们就应该尽量避免解释数据的由来？类似地，父母与子女之间在人力资本方面的相似性可能是因为基因或文化上的遗传，后者包括学校教育和人力资本方面的投资使子女获得了良好的技能与举止，从而在劳动力市场上更加有优势。跟 Becker 开创的以及 Graw 和 Mulligan 提出的模型不同，我们在分析父母与子女行为的时候所用的方法带有更多的诊断性，而不是给出产生传递过程的充分理

由。我们是揭示了从哪个角度来发现原因。这篇文章的下一部分将会探讨这个分解式。

3 认知技能的基因遗传

有一个途径应该得到充分重视，不仅仅因为乍一看它是可信的，还因为大众对这个问题的讨论给予了特别的关注。这个途径就是认知技能的基因遗传。有许多文献表明了父母与子女在认知技能上的相似性。父母与子女之间的相关系数在 0.42 到 0.72 之间，其中较高的数字是用父辈的平均智商相对于子辈的平均智商来测度的（Bouchard and McGue，1981；Plomin，DeFries，McClearn and McGuffin，2000）。通过直接或教育所获，认知机能对所得的贡献在许多使用智商（或者相关的）测试分数来估算所得决定因素的研究中得到了证实。智商对于所得的直接作用是用多元回归来分析的。回归分析中把所得的对数值作为因变量，并估算出一系列解释变量的相关系数，这些解释变量包括认知能力的测试成绩、受教育年限（也有可能是其他测度办法）、父母经济和／或社会地位的测量、工作经历、种族和性别。智商对获得较高水平的教育的贡献是通过测度孩子的智商（和其他的变量一起）而预测教育获得水平来估计的。

我们在一个所得方程（earning equation）中确定了 65 个正态回归系数估计值，这些估计值是在美国跨度达 30 年的 24 个不同研究中的数据。我们（Bowles，Gintis and Osborne，2002a）给出了荟萃分析（meta-analysis）。[3] 估计量的均值是 0.15，研究显示了在认知分数上标准离差发生一个单位的变化而其他变量（包括学校教育）保持不变时，所得的对数值产生了七分之一单位的标准离差变化。与此相比较，在用来预测所得的自然对数的同一个方程中，关于教育年限的正态回归相关系数的均值是 0.22，这表明了学校教育具有较大的独立性。

我们想通过检查确定这些结果是否依赖于作者们赋予的权重，是否会因为采取不同类型的认知测试方法和在什么年龄进行测试以及在研究中其他的一些不同而产生差异，结果发现没有很显著的区别。关于孩童时期智商对于以后学校教育（也是正态的）的影响，有研究得出一个估计值是 0.53 (Winship and Korenman, 1999)。综上所述，我们可以得出一个用来测算智商对于所得的直接和间接效应的粗略计算公式，如果用 b 来表示的话，就是 $b = 0.15 + 0.53 \times 0.22 = 0.266$。

这两个事实（一是父母与子女之间在智商方面的相似性，二是智商在直接和间接影响所得中发挥着重要作用）是否暗示着认知能力的基因遗传在代际间经济地位的传递中起着重大作用？用来说明这个问题的一个方法就是在遗传传递是惟一的源泉的条件下，父母和子女间的智商有多大的相似性。同时，还要知道在没有其他传播途径的条件下，父母与子女之间的收入有多大的相似性。

要做到这些，我们需要一些遗传学的知识（有关细节参照附录和 Bowles 与 Gintis [2001] 的文章），还要了解一些专业术语如显型（表现型）、基因型、遗传可能性和遗传相关性，这些术语对于许多经济学家来说是陌生的。人的智商——即测试分数——是一种显型，而影响智商的基因是人的基因型智商（*genotypic* IQ）。遗传可能性是指上述两者的关系。假定在一个给定的环境下，基因型的一个标准离差差异是与 h 分之一的智商标准离差差异相联系，那么 h^2 表示的是智商的遗传可能性。h^2 的估计量是通过双胞胎、兄弟姐妹、姑表亲和其他在基因上有不同程度联系的人们之间智商上的相似度来估算的。它的值不可能高于 1，最近的很多研究表明估计值一般都比较低，估计值很有可能等于一半（0.5）或者小于一半（Devlin, Daniels and Roeder, 1997; Feldman, Otto and Christiansen, 2000; Plomin, 1999）。遗传相关性在数理上度量了父母与子女之间基因的相关程度，如果父母的基因是不相关的（随机婚配）的，遗传相关性为 0.5。但是人们在选择配偶时会倾向于智商上更相似的（同型交配），这种相

似性表现为它们基因的相关系数 m。这种效应使父母与子女之间的遗传相关性增加到 $(1+m)/2$。

使用以前分解式的方法，父母与子女之间智商的相关性（用 γ 表示）等于智商的基因遗传可能性乘以遗传相关性。因此我们得到 $\gamma = h^2(1+m)/2$。由智商基因遗传导致的父母和子女收入的相关性就是 γ 乘以智商对于父母收入的标准化影响，再乘以智商对于子女收入的类似影响，也就是 γb^2。另一种算法是，我们观察到智商的基因遗传是在父母收入和子女智商的相关性中发生作用的惟一途径（我们用 γb 来表明），并把此式乘以子女智商对所得的影响 b，也能导出同样的结果。

使用上面估算的值，我们计算出智商的基因遗传对于代际间收入转移的贡献为 $(h^2(1+m)/2)(0.266)^2 = 0.035(1+m)h^2$。如果智商的遗传可能性为0.5，未知的相关关系 m 为0.2（都是合理的，如果是大致估计），假定智商的基因遗传在代际收入转移中是仅有的作用机制，那么代际相关系数为0.01，或者说是观察到的代际相关系数为0.02。我们得出的结论是智商的基因遗传在代际收入转移中的影响是可以忽略不计的。注意我们的结论不会因同型交配和遗传性的假定的改变而改变。智商不是决定经济上获得成功的决定性因素。

智商基因遗传在父母收入与子女收入之间的相似性的微小贡献可能是因为在测算认知能力时所采用的方法误差所造成的。这里有两个争论：第一，测试可靠性（reliability）体现在什么地方？不管测试什么，采用的方法能够很好地测度吗？第二，测试的有效性（validity）体现在什么地方？测试方法是否测度了正确的事情？认为测试过分渲染的想法是错误的。事实上，测试是在标准所得方程中使用一些可靠的变量而进行的。（可靠性由测试与再测试之间的相关性，在测试中奇数项目与偶数项目之间的相关性，以及别的更加复杂的方法来测度。）例如，对于通常使用的 Armed Forces Qualification Test（AFQT）方法——一种用来测试职业成功的方法但现在经常用来测试认知技

能——同一个人连续两天测试分数的相关性要高于同一个人连续两天报告教育年限或收入的相关性。

第二个问题是有关测试可能测量了错误对象，它更加重要但不容易有把握地解决。是不是存在没有用现有测量工具测度的认知技能，而它不但具有高度遗传性，还会对收入所得有很大影响，因此就可能在很大程度上解释转移过程？对一般认知能力的测试方法的探索开始于 Edward Thorndike 在 1919 年发表的论文“社会智能”。这些测试方法不考虑智商，并且预测孩子成年后的成功可能。其他一些可选择的测试办法，如 Robert Sternberg 和他的合作者所做的“实际智能”（Sternberg，Wagner，Williams and Horvath，1995；Williams and Sternberg，1995）能够预测个人在一些特殊职业上能否获得经济成功。但是，尽管实际的名声和运气可能会在特定领域对个人的成功有所帮助，Thorndike 的探究到现在还没有产生取代智商的强有力的手段，更不用说是高度遗传了。我们不能把在经济上有重要作用但又未被测量的可遗传的一般认知技能的可能性剔除出去，目前只能猜测。

确实，我们倾向于认为目前可利用的测量方法过度估计了一般认知技能对于收入所得的重要性，因为从很多方面来看，做测试就像是做工作。任何优秀的成绩都是能力和动机的复合产物，其中还包括了以下一些素质，如毅力、工作热情和其他对个人所得有独特贡献的特征。这就是为什么我们尽量避免使用测试分数这个词语来描述“认知技能”，而是用更具描述性的“认知表现”的原因。Eysenck（1994，p.9），一个认知测试中的佼佼者，这样写道：“在智商测试中解决一些低层次的问题只是一个表现的测度；在智商测试中，可能是个性而非仅仅是抽象的智力对智商产生了很大影响。一次智商测试只持续一个小时或一个小时多一点，而且疲劳、警戒、激励可能也会起作用。”因此，预测收入所得的认知测度的一些解释力反映的不是认知技能，而是有助于成功完成任务的其他个人品质。

4 基因遗传和环境遗传

尽管智商基因遗传不能解释代际间经济成功的转移过程，但还存在其他可能重要的通过基因转移的特征。确实，跟异卵双胞胎相比，同卵双胞胎在收入方面存在显著的相似性，这表明基因的作用可能是很重要的。这里，我们将用双胞胎的相似性来估计收入的基因遗传和代际转移的环境因素。

但是要注意两个问题。第一，正如下面会指出的，我们的估计值对于未观察到的参数的变化十分敏感。第二，很多时候存在着一个错误的假定，即如果一个特征的遗传是显著的，那么这个特征不会因为环境的改变而改变。身高这个例子能够很容易推翻这个谬论。以美国双胞胎作为样本对身高遗传进行的研究表明，身高遗传的可能性是很明显的（大约是 0.90，Plomin et al.，2000）。此外，这个世界上不同人种之间的高度有显著区别：居住在苏丹的丁卡人（苏丹南部的黑种部族人）的平均高度是 5 英尺 11 英寸，比挪威人和美国服役军人高一点，却比南非 Hadza 狩猎—采集者高出整整 8 英寸（Floud，Wachter and Gregory，1990）。但是 1761 年挪威征集的新兵却比今天的 Hadza 人还要矮，这表明就连可遗传性很强的特征对环境也是很敏感的。我们从由于环境变化引起的特征的小比例变化的发现中得出结论：通过改变环境来改变特征的政策需要非标准的环境条件，它不同于我们的估计所设定的环境变量。

把南非作为考察的例子，我们看到在 1993 年（纳尔逊·曼德拉成为南非总统的前一年）以前，大约有三分之二的代际收入转移是因为父子都属于同一种族，种族在当时是一个关于收入的有效预测器（Hertz，2001）。把种族这个因素加入预测子女所得的方程中，这样做的结果是减少了父辈所得对子辈所得的影响，程度大概在三分之二

以上。因为“种族”所指定的特征具有高度可遗传性，而且混种婚配的情况比较少，我们发现基因遗传在代际经济地位转移中充当着一个重要的角色。从南非这个例子中我们十分清楚地知道，在南非处于种族隔离的情况下，生理特征的基因遗传对在经济上获得成功是非常重要的，然而这种生理特征的基因遗传却是源于环境的影响。造成肤色和其他种族特征的基因遗传在经济地位转移过程中起着关键作用的原因是公共政策，而不是人的本性。这些公共政策包括对种族的界定、种族间的婚配模式和其他非白人遭受的歧视待遇。因此，转移过程中基因要素的决定性作用本身不能说明公共政策能或应该在何种程度上使竞争变得公平。

使用配对个人的数据估计遗传可能性，配对个人在不同程度上有共享的基因和环境。例如，同卵双胞胎和异卵双胞胎在成长过程中面临的环境很相似，而同卵双胞胎比异卵双胞胎在基因上有更大的相似性。在一系列很强的简化假定下（在附录里有详细解释），人们能使用有亲戚关系的两个人之间遗传和环境相似性的方差来估计特征的遗传可能性，如估计收入、受教育年限或者其他一些标准的经济变量。Taubman （1976）是第一个使用这种方法的经济学家。下面的模型在计算时，假定基因和环境会影响人力资本。人力资本能够产生所得，下面的方程也将揭示这一点。但是假定财富效应和其他对收入有贡献的变量是不受环境和基因影响的，我们下面会一一加以说明。

这里我们给出假定。第一，基因和环境有相加效应——基因和环境可能是相关的，但是“优秀基因”对所得的直接作用（用它们的回归相关系数来衡量）独立于环境。因此个人的所得可以写为：

$$所得 = h（基因）+ \beta（环境）+ 异质效应$$

第二，一对人之间的基因差异（针对异卵双胞胎）与一对人之间的环境差异不相关。举个例子说，长得漂亮的双胞胎并没有得到更多的关

爱。第三，对于异卵双胞胎和同卵双胞胎来说，影响个体发展的环境是相似的。第四，父母两人各自的所得隐形性状不相关（即他们是随机婚配的）。在给出这些假定以后，我们得出了所得的遗传可能性（h^2）是异卵双胞胎和同卵双胞胎的所得相关性差异的两倍的结论。当利用最佳数据集来测算出两者的所得相关性差异是0.2时（Björklund et al. 利用瑞典双胞胎注册登记处的数据估算得出，同时 Ashenfelter 和 Krueger（1994）利用小一点的美国双胞胎数据集估算得出），在这些假定下可以得出 h^2 估计值等于0.4。

由于父母是随机婚配的，异卵双胞胎之间的基因相关性是0.5，这其中就隐含着异卵双胞胎收入所得的相关性，因为基因因素为 $h^2/2$。事实上，观察到的双胞胎所得相关系数大于这个估计值是因为双胞胎有着类似的背景环境。因此，一旦我们知道 h^2，就能利用环境的相似性程度来估计观察到环境对观察到的所得相关性产生了多大的效应。

假定中婚配是随机的、成长环境是相同的，这是不现实的，应该予以放宽。首先，我们需要知道 m_y（父母所得的基因型的相关关系）的估计值。有关的测度办法是用潜在所得来衡量。实际所得的相关性会低估匹配程度，因为许多妇女并没有全天都在工作。分配程度对显型的作用很有可能比对基因型的作用要大得多，因为分配是基于显型而不是基于基因型。以所得作为例子，我们就能看到两者并不紧密相关。假定与父母的潜在所得相关联的基因型有一半与兄弟之间的实际收入是相似的，那么两者的相关关系为0.2。

第二，因为假定在平均上同卵双胞胎所经历的环境背景并不比异卵双胞胎所经历的环境背景更为相似，所以就应该完全用遗传相似性来解释同卵双胞胎两个人之间的收入差异比较小这一事实。但是如果同卵双胞胎比异卵双胞胎经历了更加相似的环境背景（因为同卵双胞胎长得很像），那么估计量将会过高估计遗传可能性的程度。

同卵双胞胎比异卵双胞胎和普通兄弟姐妹拥有更加相似的环境背景，这一点是很有可能的（Loehlin and Nichols，1976；Feldman et al.，

2000；Cloninger，Rice and Reich，1979；Rao，Morton，Lalouel and Lew，1982）。关于在多大程度上同卵双胞胎比异卵双胞胎经历的环境背景更为相似，这方面的估计已经相当准确，我们不可能估计得更好。估计值对有关双胞胎环境相关性的差异的敏感度可以通过假设某种程度的基因和环境的统计联系来获得，异卵双胞胎的基因相关但不相同，这使他们面临的环境的相关性比同卵双胞胎的要小。

表 1 显示了多种关于基因一环境效应的估计量。跟假设一致，当基因和环境之间的相关性增加时，同卵双胞胎之间环境的相关性随之提高，这点解释了同卵双胞胎之间收入的相似性，而收入的遗传可能性随之降低。

表 1　所得的遗传可能性的估计

基因和环境的相关性	0.00	0.50	0.70	0.80
所得的遗传可能性	0.50	0.29	0.19	0.13
正态回归相关系数：				
基因与所得	0.71	0.54	0.44	0.36
环境与所得	0.29	0.33	0.38	0.44
环境的相关关系：				
异卵双胞胎	0.70	0.70	0.70	0.70
同卵双胞胎	0.70	0.80	0.90	0.97

基因与环境之间的正态回归相关系数可以表示基因和环境之间的关系。本表中假定父母的收入决定基因的相关系数是 0.2，异卵双胞胎之间的环境相关性是 0.7。我们采用美国 Twinsburg[a] 的研究成果，取同卵双胞胎收入之间的相关性为 0.56，异卵双胞胎收入之间的相关性为 0.36，并且假定所得相关性也采用收入相关性的数值。

Björklund 等人所收集的瑞典双胞胎注册数据并不仅仅是关于双胞胎的数据，还包括了许多在不同程度上有关系的人（比如说姑表亲），他采用了 Cloninger 等（1979），Rao et al.（1982）和 Feldman et al.（2000）所开创的办法来进行估计，并允许有更多的强估计（robust estimates）。

我们把表 1 中第三栏的数值当作最合理的估计值集。由此，我们

可以推出两个非常明显的结论。第一，所得的遗传可能性是很明显的。第二，环境效应一样很大。环境与所得的正态回归相关系数是 $\beta_e = 0.38$，而之前我们的分析得出受教育年限与所得之间的正态回归相关系数是0.22，可以对两者进行比较。因此，尽管教育能够把握相关环境的重要方面，但它远不是充分的。

我们的估计量 β_e和h 里面到底隐含着什么样的代际所得相关性？为了回答这个问题，除了 h 和 β_e外，我们还需要知道父母所得与他们基因之间的相关性（我们的估计中已经隐含了这种相关性），以及父母所得与他们所经历环境之间的相关性。我们在表2的第一栏给出了估计。基因的贡献可以简单地由 h 乘以父母所得与子女后代的基因型的相关性表示，即 $h^2(1+m)/2$。类似地，环境的贡献可以由 β_e乘以父母所得和环境之间的相关性（名义上是0.74）表示。最终可以计算出代际间的所得相关性是0.4。

表2　环境、基因和财富对经济地位代际转移的贡献

	所得	收入
环境	0.28	0.20
基因	0.12	0.09
财富		0.12
代际相关性	0.40	0.41

我们将在下文讨论收入栏和财富对代际转移贡献的估计值。环境和基因栏呈现了表1第三栏的特征。

鉴于我们有关智商遗传的负面看法，基因遗传可以解释代际相关系数的三分之一的估计，这有点出人意料。环境和基因令人惊奇的重要性使我们陷入困惑。如果基因对智商的贡献并不大，而且如果环境的贡献要大于学校教育的贡献，那么造成收入在代际延续的机制是什么？我们应该回到这个谜中，不过我们会借助于数据而不是关于双胞胎的研究。

5 人力资本

因为学校教育所获在代际间是持久不变的，并且它与技能及其他能在劳动力市场获得回报的特征之间有着清晰的联系，所以认为代际经济地位转移是由于人力资本所造成的观点具有很强的合理性。我们所引用的数据已经能够计算在多大比例上，代际收入相关性是由于高收入家庭的孩子能够获得更多的教育（用年份来测度）所引起的。只要把父母收入与子女教育之间的相关性（大约是0.45）乘以所得方程中学校教育的正态回归相关系数（从我们以上的分析得出是0.22），就能得出结果大约为0.10。这个贡献很显著，特别是在教育年限与智商相互独立这一约束条件下（因为我们是从所得方程中得出估计值0.22的，方程中回归因子包括了AFQT测试和其他一些测试根据）。整个贡献测算中，包括学校教育对智商的影响，学校教育对所得的影响，以及学校教育对所得的直接影响（直接影响测度中智商是一个常数）都是0.12。

通常情况下假定，测度学校教育质量的方法一旦有了充分发展，父母经济地位对子女所得的影响只会通过认知机能和学校教育来起作用，而直接影响则会消失。但是就算采用在几年间改进的测度学校教育质量的不同方法，父母收入（或所得）对子女所得的直接影响的估计值也会发生显著改变。例如，Mulligan（1999）采用从美国青年纵向研究机构（（U.S.）National Longitudinal Study of Youth）获得的20世纪90年代早期数据，首先在没有控制任何其他因素的条件下估计了一单位父母所得对数值的变化对子女所得对数值的影响，然后他又控制了许多因素，如学校教育质量的测度方法，以及AFQT、标准教育变量和人口统计学变量，并在这样的条件下进行了重新估计。结果他发现，即使加入了一些控制因素，父母所得和子女所得之间的总的统计关系（无条件）仍保持在五分之二和二分之一之间。

利用不同数据和不同方法得出的这些结果再次肯定了黑箱之谜：代际经济地位转移的相关系数中，有一半以上不能被解释。[4]

考虑到富裕家庭的孩子要比贫困家庭的孩子更为健康这一事实（Case，Lubotsky and Paxson，2001），以及考虑到健康欠佳在以后的生活中对收入有显著影响（Smith，1999），健康很有可能在代际转移过程中发挥着极其充分的作用，因为父母收入看上去对子女健康有着很大的影响，而子女的健康既不是由父母身体健康状况也不是由父母与子女之间的基因相似性决定的。

6 财富效应

经济成功在家庭里面可以通过财富和身后物的继承转移给子孙后代。这种财富遗传的生成机制在学术界还没有得到充分重视，部分是因为没有一个充分测量其他收入决定因素的面板数据集，这个数据应该能说明第二代人在步入财富遗传已经完成的年龄时的收入情况。我们注意到惟一关于这方面的研究的就是 Menchik （1979）的文章，他利用第二代死亡时的数据来给出估计值，发现在此条件下代际财富相关性大大提高。但是当研究表明财富遗传很明显影响到富裕阶层的收入分配时，我们又怀疑这种遗传效应是否对许多家庭来说很重要，因为很少有人继承很大一笔财产。Mulligan （1997）认为不动产转移的巨大财富受制于遗传税，从 1960—1999 年的数据看，美国不动产的传递大约占死亡数 2%到 4%之间。尽管这个数字遗漏了很多在日常生活中发生的实质性的继承和转移，但是对许多家庭来说，经济地位的改变的确不太可能直接通过财富和金融资产的代际转移来完成。

因此，代际财富的延续可以至少部分说明父母与子女后代在特征上的相似性影响了财富的积累，比如对于前途的定位、个人对效率的理解、职业伦理、学校教育所得，以及对待风险的态度。人们的一些特

征影响了其财富水平，比如不太顺利的人往往都是风险规避者，他们往往对前途感到悲观，而且效率意识比较差。因为这种财富与特征之间的相关性导致了财富积累，父母和子女后代在财富方面的相似性才得以提高，而不仅仅是因为遗赠和财富转移。

不管造成父母子女之间财富相似性的源泉是什么，对于从财富中获得巨大收入的家庭来说，这种相似性在一定比例上造成了代际收入的延续。用以上所提到的分解方法，财富相似性对代际收入延续的贡献度可以由父母收入与子女财富的相关性乘以收入方程中财富的正态回归相关系数得到。我们采用从 PSID 获得的数据来说明，这些数据是根据 Charles 和 Hurst (2002)的分析得到的。在这个数据集中，父母收入与子女财富的相关性（自然值和对数值）等于 0.24。子女后代的平均年龄只有 37 岁，所以这种相关性并没有反映父母死亡时的财富遗传。为了得到一个大概的正态回归相关系数，一种办法是分析百分之一的财富变化将引起百分之几的收入变化；这个弹性的变化范围是 0（对于那些拥有少量或者没有财富的人来说）到 1（对于那些除了财富没有其他收入源泉的人来说）。利用美国人口数据计算得出的一个可信的均值是 0.20，均值的计算基于影响收入的因素是平均的。我们把均值乘以一个比率就能得到正态回归相关系数，这个比率等于财富对数值的标准差除以收入对数值的标准差，这也是从 Charles 和 Hurst (2002)那里获得的数据。计算结果表明，高收入家庭的子女普遍比较富裕这一事实在代际收入相关性中的贡献度是 0.12。

这个数值尽管很大，但它很有可能被低估了。因为计算这个数值所用的数据本身是由于上面提到的原因并不能反映财富转移过程中的关键要素，也就是财产遗传要等到父母过世才能实现。此外，如果考虑到较大的财富拥有较高的平均回报这一事实，这个估计值还应该经过调整(Bardhan, Bowles and Gintis, 2000; Yitzhaki, 1987)。较大的父母财富和自有财富还会提高学校教育和其他人力资本的回报比率，但是我们在实际工作中没有办法来考虑这些。对于一个父母富裕的家庭样

本来说，财富对代际相关性的贡献度将会更高。对于一个出生在只有少量财富的家庭样本来说，财富对代际相关性的贡献度将接近于零。财富效应在不同的收入分布中的差别反映了我们前面提到的在转移过程中的异质特征。由于财富的偏度分布，我们可以认为处于财富均值水平（我们估计中应用此均值）的家庭要比中等水平的家庭更加富裕。

7 群体成员和个性

到目前为止，我们都是采用生产函数法来分析，这种分析方法成为其他分析代际转移的经济学分析法的基础，此分析法试图计算出所有关于生产要素所有权对父母子女相似性的贡献度。在考虑到基因遗传的条件下，我们已经补充了一些常用的可供选择的分析法。但是对于其他代际延续的特征在分析中是否很重要这一点却很有争议，这些特征有种族、母语、孩子数量、兄弟姐妹数量等等。生理特征是一种具有很强预测作用的可遗传非技能特征。例如，肥胖是收入很低的妇女的特征，而高度却是男性高收入的特征。好的相貌是高收入者的特征，这一点对于男性和女性都一样。对于女性而言不管她是否拥有与公众社会打交道的职业，好的相貌对其收入来说都是很关键的（Hammermesh and Biddle，1993）。Bowles，Gintis和Osborne（2002a）的研究给出了实际证据来说明非技能特征对经济上获得成功具有决定性意义。

有两个变量能够说明非技能性因素在代际经济地位转移中的可能重要性：族群成员和个性。

假设经济成功不仅仅受个人特征的影响，还受到个人所接触的群体圈子的特征的影响。群体可以用多种维度来加以区分：学校教育平均水平，经济上的成功，认知机能和财富水平。群体可以指地域上的居民，可以是邻居，可以是伦理道德上或种族上的群体，可以是共同语言意义上的群体，可以是一个国家的公民，还可以是个人在交往中接触到

的有代表性的人的集合。有研究表明了群体对经济成功的影响（Cooper et al.，1994；Durlauf，2001；Borjas，1995），并且这种影响还会因为一些原因而加强，这些原因包括歧视、墨守成规对行为的作用、获得信息的不同途径、产品补贴等等。

很明显，种族在代际经济成功转移方面起了重大作用。Björklund，Eriksson，Jäntti，Raaum 和 Osterbacka（2002）的研究证明了这一点。他们用美国的数据估计出兄弟间所得的相关性是 0.43，但是当单独利用白人作为研究样本时估计值降到 0.10。很明显，兄弟间在很多方面都是一直相同的，那么这就说明种族是造成他们收入相似性的主要原因。用父母和子女来作分析也能得到相同的结果。在图 1 所呈现的数据集中，后代家庭收入对父母家庭收入的弹性是 0.54，但是对白人而言弹性值仅为 0.43，对黑人而言弹性值仅为 0.41（Hertz，2002）。以上“种族”在代际间转移的重要测度解释了父母与子女之间在收入方面的相似性。利用 Hertz 的估计值我们发现种族（如父母收入与子女种族的相关性）在代际相关性中的贡献度是 0.07。尽管这个估计值与前面提到的数据相比有点低，但是还是有可能被高估了，因为估计值是用收入方程的标准回归量计算得到的，而没有测度认知表现，加入认知表现就有可能稍微降低种族的相关系数。

在传统的生产函数中没有体现出第二个特征——个性，但是个性却通过个人效率意识、职业伦理或者低时间贴现率（对前途有很强的定位）这些方式对代际经济地位转移发挥着作用。这些特征在各方面的重要性可以用一系列的交易加以说明，这些交易包括劳动力雇佣、资金借贷，或者在质量不确定情况下的商品交易。交易过程中各方不可能依靠具有法律强制效力的合同来详细说明交易的各个方面，在这种条件下，交易各方只能通过彼此的互相信任程度、诚信状况、工作努力程度和其他对方所拥有的特征来保障交易的顺利完成。举个例子来说，一个只关注当前利益的雇员将不会考虑雇主对于将来续聘的承诺，他们只会在适当的条件下努力工作。在目前，雇主需要用更高的薪酬来激励

他们努力工作，因此这些雇员在劳动力市场上被雇佣的可能性也较小。再比如说，相信宿命论的工作者认为他们当前的行为不会使他们被炒鱿鱼，所以这种劳动力雇佣合约要想激励雇员努力工作将会耗费过高的成本（Bowles, Gintis and Osborne, 2002a）。Duncan 和 Dunifon (1998)，Heckman 和 Rubinstein (2001)，Kuhn 和 Weinberger (2001)，Heckman（即将出版）等人的一系列研究都表明了这些特征的现实重要性。

Osborne（即将出版）研究了宿命论的经济重要性和代际延续，就像罗德尺度（Rotter Scale）测度的一样，她采用一种普通的办法来测量在多大程度上个人认为他们一辈子中发生的重大事情都是由外在的事情引起的，而不是因为个人的努力。她的研究取样于美国男性以及这些男性的父母，她发现宿命论对美国男性在步入劳动力市场前的影响从统计上看显著，并且对所得有很大影响。此外，她发现父母与子女的 Rotter 分数都是一致的。Osborne 的研究中得到的正规影响数值（在绝对值意义上，也就是 −0.2）比我们前面讨论所用荟萃分析法得出的智商平均影响数值稍微要高一点。父母收入与子女宿命论的相关性的估计值为 −0.14。宿命论对代际相关性的贡献可以用父母收入与子女宿命论的相关性乘以子女宿命论与其以后收入之间的相关性得到，这里是 0.028，以（−0.2）×（−0.14）得到。

Osborne 还以英格兰的女性为样本进行了研究，她发现在样本 11 岁的时候测量社会失调问题（用 Bristol 社会调整尺度），比如说进取与消极，将能够有效预测样本 33 岁时候的所得。进取与消极等个人特征对所得的正规影响要比智商的影响要大得多。在 Osborne 的研究中没有测度这些英国数据集里面的代际个性特征的一致性，但是有其他的研究表明用社会失调法来测度的父母与子女之间相似性数值可能相当高。如 Duncan, Kalil, Mayer, Tepper 和 Payne（即将出版）发现美国母亲的不正常行为举止将会影响到她们的女儿，包括吸毒、暴力行为、早期性行为、中途辍学和犯罪倾向。Osborne 的研究表明了代际个性特征

的转移（无论是基因上的还是文化上的）可能是解释代际间收入一致性的重要渠道。

相对于认知机能，我们对与经济成功相关的个性特征的代际转移过程知之甚少。然而 Melvin Kohn（1969）关于父母抚养孩子的价值观的研究表明，至少有一些特征如父母在职场的经历可以总结起来传递给子女。Kohn 以父母在工作中的自主程度对父母的样本进行分类，从无人监督型到完全受主管监控。Kohn 发现处在较高水平的父母——他称之为“职业的自我管理”者（occupational self-direction）——将教导他们的子女树立好奇心、自我控制、幸福和独立的价值观念。而那些在工作中受到主管紧密监视的父母将会强调服从外在的权威。Kohn 得出结论：“无论是否有意识，父母总是倾向于以他们自身所处的社会阶层所得到的经验教训来教导他们的子女，因此这也为他们的子女进入类似的社会阶级做了准备。”Osborne 所作的研究提出自我管理程度对后代的所得有显著影响。其他一些研究如 Yeung，Hill 和 Duncan（2000）表明父母行为，包括宗教参与、社会组织成员关系和一些预防性的行为如乘车时系安全带，都会对子女的所得有很大的影响。

8 结　论

最近的证据表明代际经济地位转移的程度水平要比先前预想得高。从别的一些测量结果来看，美国依然是一个充满机会的国度，但是父母的收入和财富却明确预示了下一代将很有可能处于什么样的经济地位。

我们的目标是估计出代际转移的程度，以及揭示转移的生成机制是什么。表 3 给出了关于我们能够识别的各种主要生成渠道的相对重要性的最优估计值。表中惟一在之前没有给出解释的就是第一行，第一行是受学校教育年限和其他一些条件下的智商，它等于父母收入与子女智商之间相关性的估计值乘以智商对所得的正规影响的估计值。

在收入栏中关于智商、学校教育和个人所得的估计值是经过了简单调整后的所得栏的估计值，调整是考虑到所得差异对收入差异的影响后进行的修正，估计值和鲍尔斯与金迪斯（2001）的估计值很相符。因此，我们不考虑这些所得的决定因素可能会影响到个人财富的回报率。相反，我们假定种族对人力资本和传统财富的报告有着相同的作用（如果种族仅仅通过对所得的影响来影响收入，它对代际间所得的相关性将会明显变得更大）。

表 3　美国经济地位代际转移的主要渠道

渠道	所得	收入
智商，在学校教育条件下	0.05	0.04
学校教育，在智商的条件下	0.10	0.07
财富		0.12
个性特征（宿命论）	0.03	0.02
种族	0.07	0.07
总代际相关性估计	0.25	0.32

对每一种渠道而言，所列示的数值由父母收入与预测到的子女收入的相关性，乘以它们在所得方程或收入方程中的正态回归相关系数得到。总代际相关性是由各种渠道加总得到，其中不包括父母经济地位对子女经济地位直接影响。

数据来源：由正文中描述的公式计算以及鲍尔斯和金迪斯（2001）的文章。

尽管表 3 的估计值相当不精确，但结果的性质却不太可能因为选用其他一些可供选择的方法而改变。这有点令人惊讶，财富、种族和学校教育对经济地位的转移很重要，但智商不是主要贡献因素，并且就像我们前面看到的，智商遗传传递的作用更小。

关注经济地位代际转移的政策制订者会遇到两个不同的难点。第一，许多可能会影响到经济地位代际转移的政策常常引发争议。例如，在目前的政治气候下，以增加不动产税的办法来限制代际财产转移通常被认为是不友善的方式。消除种族歧视虽然是减少收入遗传可能性的方式之一，但是要达到这个目标十分困难。通过改进教育成就的方式，特别是对于那些其父母受学校教育水平较低的孩子来说，将会直

接和间接地减少经济地位代际转移的可能性。直接作用是指学校教育的影响，间接作用是指提供了一个更加开放的群体关系的网络结构和婚配选择，而不仅仅是用收入来界定同类关系。但是改进教育成就也是一个很难达到的目标。

第二个难点是规范。公平竞争的环境能保证父母收入和子女的收入之间没有相关性吗？（Suift，即将出版）家庭生活中的价值观念和隐私可能会因为任何尝试拆离父母与子女财富的努力而不得不折中考虑。与其追求一个零代际相关性的抽象（对于我们而言是不吸引人的）目标，不如动手处理一个更好的问题，即了解哪种代际转移生成机制是不公正的，并以此来制定相关政策。种族在地位代际转移中所发挥的作用无疑是不公正的。许多人认为父母收入与子女健康有很强的相关性的说法在道德上很值得怀疑，另外许多人也从道德上对财富遗传的高水平程度表示怀疑。大多数人赞同对先天性残疾人进行补偿的政策。其他关于父母子女之间收入延续的作用机制，例如美丽相貌的基因遗传，被认为是不会招致反对意见的，所以不是政策所要干预的恰当目标。尽管有些一致意见能够建立在道德上令人怀疑的机制之上，但是这些政策的含意远未清楚。例如，我们不得不考虑和估计父母的一些可能动机，这些动机使父母对子女成功的影响减少。

解决政策挑战不仅需要对上面所说到的难点的明确道德性及相关问题有一个清楚理解，更需要了解是哪一种作用机制造成了实质性的代际一致性经济差别。

9 附录：相关系数分解式和遗传可能性估计

假设父母收入 y_p 直接影响子女后代所得 y，子女所得同时还受到两个变量 v_1 和 v_2 的影响，这两个变量与父母所得相关联。[5] 如果分别用 $r_{y_pv_1}$ 和 $r_{y_pv_2}$ 表示父母所得与 v_1 和 v_2 的相关性，用 β_{y_py}，

β_{v_1y}和 β_{v_2y}分别来表示 y_p， v_1， v_2 与预测值 y 的正态回归相关系数，那么我们有：

$$r_{y_py} = \beta_{y_py} + r_{y_pv_1}\beta_{v_1y} + r_{y_pv_2}\beta_{v_2y} \tag{1a}$$

父母所得与子女所得之间的相关性可以分解成直接影响（第一项）、由变量 v_1 产生的影响（第二项）和由变量 v_2 产生的影响（第三项），方程可以表示为：

$$y = \beta_{y_py}y_p + \beta_{v_1y}v_1 + \beta_{v_2y}v_2 + \varepsilon_y \tag{2a}$$

这里所有变量都是服从均值为 0，方差为 1 的正态分布， ε_y与独立变量不相关。那么，用期望值 $E[y_py]$ 代替 y，重写方程式，由于两个变量（如 y 和y_p）的均值为 0，方差为 1，故这些变量之间的相关性等于它们乘积的期望值，则方程为：

$$r_{y_py}y = E[y_py] = E[r_{y_py_p}]\beta_{y_py} + E[v_1y]\beta_{v_1y} + E[v_2y]\beta_{v_2y} \tag{3a}$$

所以，进行正规化，$E[r_{y_py_p}] = 1$，$E[v_1y] = r_{v_1y}$，$E[v_2y] = r_{v_2y}$，我们就得到了（1a）式。[6]

我们现在应用这种办法对同卵双胞胎和异卵双胞胎相似性方面的数据进行估计。一种更为一般的办法是利用两个人之间的不同关联程度来进行估计，这种办法是由 Feldman et al.（2000）开创的。假设同一个家庭里的两个儿子的所得 y_1 和 y_2，分别由他们的基因型 g_1 和 g_2 以及他们所处的环境 e_1 和 e_2 决定，那么：

$$y_i = \beta_e e_i + hg_i + \varepsilon_{y_i} \quad 对于\ i = 1，2 \tag{4a}$$

这里 ε_{y_i}跟模型里的独立变量不相关，并且 y_i 的方差是单一的。e_i和g_i的方差也是相同的。注意，这里基因的正态回归相关系数 h 是所得遗传可能性的平方根。我们假定兄弟 i 的环境变量e_i 决定于他自身基因g_i 和共同背景E。因此我们有：

$$e_i = \beta_E E + \beta_{ge}g_i + \varepsilon_{e_i} \quad 对于\ i = 1，2 \tag{5a}$$

这里 ε_{e_i} 跟模型里的独立变量不相关，并且 e_i 的方差是单一的。共同环境 E 包括了父母所得的影响、教育，以及任何兄弟间共同享有的能够影响子女后代所得的其他因素。为简单起见，我们用相关系数 β_{ge} 表示基因对环境的全部影响，故 g_i 与共同环境 E 不相关。最后，兄弟 i 的基因 g_i 是由父母的基因决定的，用方程表示为：

$$g_i = \beta_g g_f + \beta_g g_m \tag{6a}$$

这里 g_f 是指父亲的基因，g_m 是指母亲的基因，β_g 是用父亲或母亲的基因来预测儿子基因时的正态回归系数。这个模型的结构可以用图 3 来说明。

为了给出 β_g 为 1/2，假设 m_y 为母系基因和父系基因的相关系数。既然已经假定了可加性（即基因组的总影响等于各个基因影响的加总），我们就可以用单一的轨迹来表示 β_g。在这条轨迹中，我们用 x 来表示可能对所得有贡献的基因的总和。正规化 x，我们得到 $E[x] = 0$ 以及 $(x_f + x_m)/2$。根据基本遗传学知识，在这条轨迹中儿子从父母亲身上各获得一份基因，x_f 表示从父亲身上获得的基因，x_m 表示从母亲身上获得的基因。那么在这条轨迹中儿子的基因值就等于 $(x_f + x_m)/2$，这里我们假定两种基因对经济成功都有相同的影响。在全文中我们都遵循这个假定。[7] 在这条轨迹中，除了 x_f，父亲还有另一种基因值 z_f，同样也是均值为 0，方差为 2。父系基因相应的值为 $(z_f + x_f)/2$，其中 x_f 与 z_f 不相关。母系基因相应的值为 $(z_m + x_m)/2$，其中在这条轨迹中 z_m 是母系的其他基因，并且 x_m 与 z_m 不相关。因为父母的婚姻很匹配，每一个父系基因 x_f 和 z_f 以 m_y 与每一个母系基因相关。在这条轨迹中父母亲基因值的方差为：

$$E[(x_m + z_m)^2/4] = E[(x_f + z_f)^2/4] = 1$$

父子之间基因的协方差为：

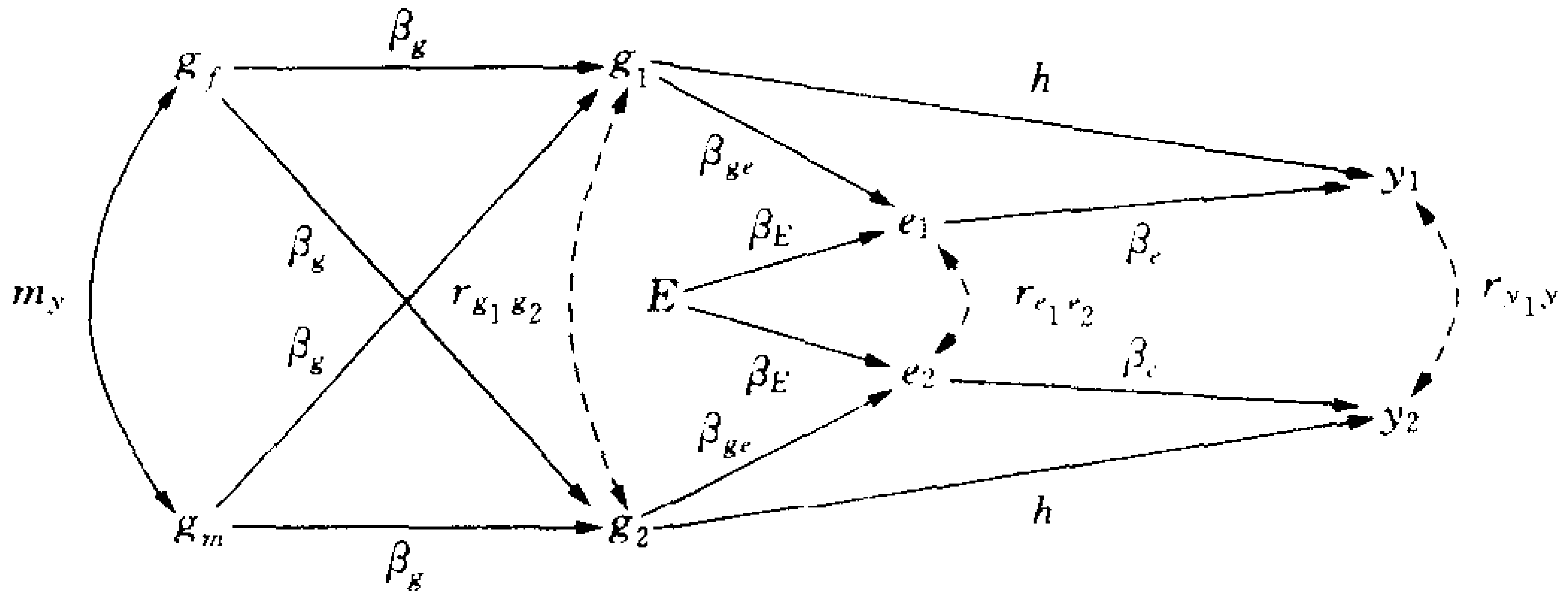

图3　兄弟所得

本图中，g_f 是指父亲的基因，g_m 是指母亲的基因，g_1，g_2 分别是指兄弟的基因，E 是指兄弟所拥有的共同环境，e_1、e_2 分别指兄弟所经历的总的环境背景，y_1，y_2 分别是指兄弟的所得。这里 m_y 是指婚姻匹配的父母的基因联系，下面我们会给出具体解释，$\beta_g=1/2$，h^2 是指所得的遗传可能性。路径 β_{ge} 表示基因影响环境的趋势（$\beta_{ge}>0$ 意味着同卵双胞胎所经历的环境比异卵双胞胎经历的环境更为相似）。

$$E[(x_f+z_f)(x_f+x_m)/4]=(1+m_y)/2$$

因此在这条轨迹中，父系基因与儿子基因的相关性等于前面两个表达式的商，即：

$$r_{g_fg_i}=\beta_{ge}+\beta_{ge}m_y=\frac{1}{2}+\frac{m_y}{2} \tag{7a}$$

这个表达式中的第一项代表从父系基因组遗传到儿子基因组的直接路径，第二项代表这条轨迹图中父系基因值和母系基因值的相关性，即 m_y，乘以这条轨迹中母亲传给儿子的直接路径。为了得到该式，我们回忆最小二乘回归公式：$y=b_1x_1+b_2x_2+\varepsilon$，这里 x_1，x_2，y 服从均值为0、方差相同的正态分布，ε 与 x_1，x_2 不相关，则 Goldberger（1991）给出了：

$$b_1=\frac{r_{x_1y}-r_{x_1x_2}r_{x_2y}}{1-r_{x_1x_2}}$$

在我们的例子中，$b_1=\beta_g$，$r_{x_1y}=r_{x_2y}=(1+m_y)/2$ 并且 $r_{x_1x_2}=m_y$。代入上式，我们得到 $\beta_g=1/2$。

为了确定异卵双胞胎之间基因的相关性，我们取 $i=1$，2，并乘以（6a）式中的右边部分，然后取期望值，则：

$$r^{fr}_{g_1g_2}=E\ [g_1g_2]\ =\ (1/2)^2E\ [g_f^2]$$
$$+\ (1/2)^2E\ [g_m^2]\ +2\ (1/2)^2E\ [g_mg_f]$$
$$=\ (1/2)^2\ (2+2m_y)\ =\ (1+m_y)\ /2$$

比照（7a）式，我们得到标准结果，发现父子间基因和具有相同父母的非同源兄弟间基因有相等的联系。为了确定异卵双胞胎所经历环境之间的相关性，我们取 $i=1$，2，并乘以（5a）式中的右边部分，然后取期望值，则：

$$r^{fr}_{e_1e_2}=\beta^2_E+r^{fr}_{g_1g_2}\ \beta^2_{ge}$$
$$=\beta^2_E+\ (1+m_y)\ \beta^2_{ge}/2$$

最后，我们取 $i=1$，2，并乘以（4a）式中的右边部分，然后取期望值，则：

$$r^{fr}_{y_1y_2}=\beta^2_e r^{fr}_{e_1e_2}+h^2r^{fr}_{g_1g_2}+2\,\beta_e h r_{g_1g_2}\ \beta_{ge}$$

上式还可以扩展为：

$$r^{fr}_{y_1y_2}=\beta^2_e\ (\beta^2_E+\ (1+m_y)\ \beta^2_{ge}/2)\ +h^2\ (1+m_y)\ /2$$
$$+\ (1+m_y)\ \beta_e\beta_{ge}h \qquad (8a)$$

我们采用同样的办法和同样的有关数字来计算同卵双胞胎之间所得的相关性，但是现在兄弟间基因型的相关性 $r^{id}_{g_1g_2}=1$，那么

$$r^{id}_{e_1e_2}=\beta^2_E+r^{id}_{g_1g_2}\ \beta^2_e$$
$$=\beta^2_E+\ \beta^2_{ge}$$

则

$$r^{id}_{y_1y_2}=\beta^2_e r^{id}_{e_1e_2}+h^2r^{id}_{g_1g_2}+\ \beta_e h r^{id}_{e_1g_2}+\ \beta_e h r^{id}_{e_2g_1}$$

可以变成：

$$r^{id}_{y_1y_2}=\beta^2_e\ (\beta^2_E+\ \beta^2_{ge})\ +h^2+2\,\beta_e\beta_{ge}h \qquad (9a)$$

在正文中，我们假定同卵双胞胎的 $r^{id}_{e_1 e_2} = 0.9$（尽管我们的结果对这个假定不是十分敏感），所以 $\beta_E = \sqrt{0.9 - \beta^2_g}$。两个关于兄弟所得之间相关性的方程（8a）和（9a）和由这些相关性观察到的值，使我们能够根据多个 β_g 值来确定 h 和 β_e 的值。方程（8a）和（9a）隐含着同卵双胞胎之间所得相关性与异卵双胞胎之间所得相关性的差异，这个差异可以由下式给出：

$$r^{id}_{y_1 y_2} - r^{fr}_{y_1 y_2} = (1 - m_y)(h + \beta_e \beta_{ge})^2 / 2 \qquad (10a)$$

注意，我们假定更加匹配的婚姻会提高 h^2 的估计值，而假定基因影响环境的趋势越强（β_{ge} 值提高），则 h^2 的估计值越低。在文献中，通常假定 $m_y = 0$ 和 $\beta_{ge} = 0$，因此在这个例子中，我们就能得到估计遗传可能性的标准方程：

$$h^2 = 2(r^{id}_{y_1 y_2} - r^{fr}_{y_1 y_2}) \qquad (11a)$$

如果事实真是这样，我们就能够直接从这个方程中估计 h^2，然后用 h^2 的估计值联合方程（8a），就能估计出 β_e。

注释：

[1] 参见鲍尔斯和金迪斯(2001)所作的有关正式模型和其他关于技术方面的研究。Arrow，Bowles 和 Durlauf（1999），以及鲍尔斯，金迪斯(eds.，forthcoming)的研究都给出了最近一些的实证和理论成果。

[2] 此分解式能够在 Blalock（1964）中找到，在本文的附录中也有详细介绍。Goldberger（1991)用正态（均值为 0，标准差为 1）变量来说明标准回归模型。

[3] 该方法是由 Hunter 和 Schmit 所创造的一种统计分析方法，它使研究者能矫正对同一个问题的各项统计研究中的人为因素的影响，从而获得两个研究变量之间的真实联系。——译者注

[4] 利用以上所描述的传统变量，我们也可以说能够在统计上解释不到一半的所得或收入差异。但是这个事实并不能解释我们在代际相关性方面的有限成功，因为这个相关性仅仅测度了所得的部分变化，这部分变化在统计上我们能够用父母经济地位来解释。

[5] 具体处理办法，参见 Rao，Morton Yee（1976），Cloninger et al.（1979），Rao et al.（1982），以及 Otto，Feldman 和 Christiansen（1994）的文章。

[6] 注意，如果我们用样本均值，样本方差和样本协方差来替换这些总人数的期望值，就会出现相同的争论。在这个例子中，误差项和独立变量的统计独立是在构建模型时就确立的，而在总人口水平上这种独立性却是被假定的。

[7] 当然，在轨迹图中，一对基因的实际值有可能比平均值要高一点或低一点，因为有的基因是显性的，有的基因是隐性的。

参考文献:

Arrow, Kenneth, Samuel bowles, and Steven Durlauf, *Meritocracy and Economic Inequality* (Princeton, NJ: Princeton University Press, 1999).

Ashenfelter, Orley and Alan Krueger, "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review* 84, 5 (December 1994): 1157—1172.

Atkinson, A. B., A. K. Maynard, and C. G. Trinder, *Parents and Children: Incomes in Two Generations* (London: Heinemann, 1983).

Bardhan, Pranab, Samuel bowles, and Herbert gintis, "Wealth Inequality, Credit Constraints, and Economic Performance," in Anthony Atkinson and François Bourguignon (eds.) *Handbook of Income Distribution* (Dortrecht: North-Holland, 2000).

Becker, Gary S., "Family Economics and Macro Behavior," *American Economic Review* 78 (1988): 1—13.

— and Nigel Tomes, "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics* 4, 3 (July 1986): S1—39.

Björklund, Anders, Markus Jäntti, and Gary Solon, "Influences of Nature and Nurture on Earnings: An Early Progress Report on a Study of Various Sibling Types in Sweden," in Samuel bowles, Herbert gintis, and Melissa Osborne (eds.) *Unequal Chances: Family Background and Economic Success* (New York: Russell Sage Foundation, forthcoming).

—, Tor Eriksson, Markus Jäntti, Oddbjorn Raaum, and Eva Osterbacka, "Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States," *Journal of Population Economics* (2002).

Blalock, Hubert, *Causal Inferences in Nonexperimental Research* (1964).

Blau, Peter and Otis Dudley Duncan, *The American Occupational Structure* (New York: Wiley, 1967).

Borjas, George, "Ethnicity, Neighborhoods, and Human Capital Exernalities," *American Economic Review* 85, 3 (June 1995): 365—390.

Bouchard, T. J., Jr. and M. McGue, "Familial Studies of Intelligence," *Science* 212 (1981): 1055—1059.

Bowles, Samuel, "Schooling and Inequality from Generation to Generation," *Journal of Political Economy* 80, 3 (May/June 1972): S219—251.

— and Herbert Gintis, "The Inheritance of Economic Status: Education, Class and Genetics," in Marcus Feldman and Paul Baltes (eds.) *International Encyclopedia of the Social and Behavioral Sciences: Genetics, Behavior and Society* (New York: Oxford University Press and Elsevier, 2001).

— and Valerie Nelson, "The 'Inheritance of IQ' and the Intergenerational Reproduction of Economic Inequality," *Review of Economics and Statistics* 56, 1 (February 1974).

—, Christina Fong, and Herbert Gintis, "Reciprocity and the Welfare State," in Jean Mercier-Ythier, SergeKolm, and Louis-André G'rard-Varet (eds.) *Handbook on the Economics of Giving, Reciprocity and Altruism* (Amsterdam: Elsevier, 2002).

—, Herbert Gintis, and Melissa Osborne, "The Determinants of Individual Earnings: Skills, Preferences, and Schooling," *Journal of Economic Literature* (2002): 1137—1176.

— and Melissa Osborne (eds.), *Unequal Chances: Family Background and Economic Success* (New York: Russell Sage Foundation, forthcoming).

Case, Anne, Darren Lubotsky, and Christina Paxson, "Economic Status and Health in Childhood: The Origins of the Gradient," June 2001. NBER Working Paper No. W8344.

Charles, Kerwin Kofi and Erik Hurst, "The Correlation of Wealth Across

Generations," 2002. University of Chicago Working Paper.

Cloninger, C. Robert, John Rice, and Theodore Reich, "Multifactorial Inheritance with Cultural Transmission and Assortative Mating. II. A General Model of Combined Polygenic and Cultural Inheritance," *American Journal of Human Genetics* 31 (1979): 176—198.

Cooper, Suzanne, Steven Durlauf, and Paul Johnson, "On the Evolution of Economic Status across Generations," *American Economic Review* 84, 2 (1994): 50—58.

Corak, Miles and Andrew Heisz, "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from the Longitudinal Income Tax Data," *Journal of Human Resources* 34, 3 (1999): 505—533.

Devlin, Bernie, Michael Daniels, and Kathryn Roeder, "The Heritability of IQ," *Nature* 388 (31 July 1997): 468—471.

Duncan, Greg, Ariel Kalil, Susan E. Mayer, Robin Tepper, and Monique R. Payne, "The Apple Does Not Fall Far from the Tree," in Samuel Bowles, Herbert Gintis, and Melissa Osborne (eds.) *Unequal Chances: Family Background and Economic Success* (New York: Russell Sage Foundation, forthcoming).

Duncan, Greg J. and Rachel Dunifon, " 'Soft-Skills' and Long-Run Market Success," *Research in Labor Economics* 17 (1998): 123—150.

Durlauf, Stephen, "A Framework for the Study of Individual Behavior and Social Interactions," *Sociological Methodology* 31 (2001): 123—128.

Erikson, R. and J. H. Goldthorpe, *The Constant Flux: a Study of Class Mobility in the Industrial Societies* (Oxford: Oxford University Press, 1992).

Eysenck, H. J., "Intelligence and Introversion-extraversion," in Robert J. Sternberg and Patricia Ruzgis (eds.) *Personality and Intellegence* (Cambridge University Press, 1994) pp. 3—31.

Feldman, Marcus W., Sarah P. Otto, and Freddy B. Christiansen, "Genes, Culture, and Inequality," in Kenneth Arrow, Samuel Bowles, and Steven Durlauf (eds.) *Meritocracy and Economic Inequality* (Princeton University Press, 2000) pp. 61—85.

Floud, Roderick, Kenneth Wachter, and Annabel Gregory, *Height, Health and History: Nutritional Status in the United Kingdom, 1750—1980* (Cambridge: Cambridge University Press, 1990).

Fong, Christina, "Social Preferences, Self-Interest, and the Demand for Redistribution," *Journal of Public Economics* 82, 2 (2001): 225—246.

Goldberger, Arthur S., "Economic and Mechanical Models of Intergenerational Transmission," *American Economic Review* 79, 3 (June 1989): 504—513.

—, *A Course in Econometrics* (Cambridge, MA: Harvard University Press, 1991).

Hammermesh, Daniel S. and Jeff E. Biddle, "Beauty and the Labor Markets," November 1993. NBER Working Paper 4518.

Heckman, James, *The GED* (forthcoming). and Yona Rubinstein, "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," *American Economic Review* 91, 2 (May 2001): 145—149.

Hertz, Thomas, "Education, Inequality and Economic Mobility in South Africa," 2001. Unpublished Ph. D. Disseration, University of Massachusetts. "Intergenerational Economic Mobility of Black and White Families in the United States," May 2002. Paper presented at the Society of Labor Economists, Annual Meeting.

Kohn, Melvin, *Class and Conformity* (Homewood, IL: Dorsey Press, 1969).

Kuhn, Peter and Catherine Weinberger, "Leadership Skills and Wages," December 2001. Institute for Social, Behavioral and Economic Research.

Loehlin, John and Robert Nichols, *Heredity, Environment, and Personality* (Austin, TX: University of Texas Press, 1976).

Mazumder, Bhashkar, "Earnings Mobility in the U. S.: A New Look at Intergenerational Inequality," in Samuel Bowles, Herbert Gintis, and Melissa Osborne (eds.) *Unequal Chances: Family Background and Economic Success* (New York: Russell Sage Foundation, forthcoming).

Menchik, Paul, "Intergenerational Transmission of Inequality: An Empirical Study of Wealth Mobility," *Economica* 46 (1979): 349—362.

Mulligan, Casey, *Parental Priorities and Economic Inequality* (Chicago: University of Chicago Press, 1997).

— "Galton vs. the Human Capital Approach to Inheritance," *Journal of Political Economy* 107, 6 (December 1999): S184—224. Part 2.

Osborne, Melissa A., "Personality and the Intergenerational Transmission of Economic Status," in Samuel Bowles, Herbert Gintis, and Melissa Osborne (eds.) *Unequal Chances: Family Background and Economic Success* (New York: Russell Sage Foundation, forthcoming).

Otto, Sarah P., Marcus W. Feldman, and Freddy B. Christiansen, "Genetic and Cultural Transmission of Continuous Traits," 1994. Stanford University, Morrisson Institute Working Paper No. 64.

Plomin, Robert, "Genetic and General Cognitive Ability," *Nature* 402, 2 (December 1999): c25—c29.

—, John C. DeFries, Gerald McClearn, and Michael McGuffin, *Behavioral Genetics* (New York: W. H. Freeman and Company, 2000).

Rao, D. C., N. E. Morton, and S. Yee, "Resolution of Cultural and Biological Inheritance by Path Analysis," *American Journal of Human Genetics* 28 (1976): 228—242.

—, —, J. M. Lalouel, and R. Lew, "Path Analysis Under Generalized Assortative Mating. II: American IQ," *Genetic Research*, Cambridge 39 (1982): 187—198.

Smith, James, "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status," *Journal of Economic Perspectives* 13, 2 (1999): 145—166.

Solon, Gary R., "Intergenerational Income Mobility in the United States," *American Economic Review* 82, 3 (June 1992): 393—408.

—, "Intergenerational Mobility in the Labor Market," in Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics* (Amsterdam: North-Holland, 2000).

Sternberg, Robert J., Richard K. Wagner, Wendy M. Williams, and Joseph Horvath, "Testing Common Sense," *American Psychologist* 50, 11 (November 1995): 912—927.

Swift, Adam, "Justice, Luck and Family Values: The Intergenerational Transmission of Economic Status from a Normative Perspective," in Samuel Bowles, Herbert Gintis, and Melissa Osborne (eds.) *Unequal Chances: Family Background and Economic Success* (New York: Russell Sage Foundation, forthcoming).

Taubman, Paul, "The Determinants of Earnings: Genetic, Family, and Other Environments; A Study of White Male Twins," *American Economic Review* 66, 5 (December 1976): 858—870.

Thorndike, Edward L., "Intelligence and Its Uses," *Harper's Monthly Magazine* 140 (December/January 1919): 227—235.

Williams, Wendy M. and Robert J. Sternberg, *Success Acts for Managers* (Florida: Harcourt Brace, 1995).

Winship, Christopher and Sanders Korenman, "Economic Success and the Evolution of Schooling with Mental Ability," in Susan Mayer and Paul Peterson (eds.) *Earning and Learning: How Schools Matter* (Washington, DC: Brookings Insitution, 1999) pp. 49—78.

Yeung, Jean, Martha Hill, and Greg Duncan, "Putting Fathers Back in the Picture: Parental Activities and Children's Adult Attainments," *Marriage and Family Review* 29, 2/3, Part I (2000): 97—113.

Yitzhaki, S., "The Relation Between Return and Income," *Quarterly Journal of Economics* (February 1987): 77—95.

Zimmerman, David J., "Regression Toward Mediocrity in Economic Stature," *American Economic Review* 82, 3 (June 1992): 409—429.

人类利他行为的解释*

赫伯特·金迪斯　萨缪·鲍尔斯
罗伯特·博依德　恩斯特·费尔

1 引　　言

包括适存度理论和互惠利他主义（Hamilton，1964；Trivers，1971；Williams，1966）的解释使一代研究者确信，所谓的利他主义——为了他人牺牲自己的利益——只不过是从长期来说的自利而已。例如道金斯（1976，1989）在《自私的基因》里提出了这样的看法。他自信地断言："我们是幸存的机器——自动的机械盲目地设置和保存了一种叫做'基因'的自私分子……正是这个自私的基因导致了人类的自私行为。"道金斯虽然考虑到了社会生活中的道德，但这个道德必须强加在一个本质自私的人身上。"我们能做的只是尽最大可能来宣扬慷慨和利他，""因为我们天生是自私的。"然而，根据William — Hamilton传统中最有影响力的伦理学家R. D. Alexander的理解，甚至社会道德也

* 原文题目为 Explaining altruistic behavior in humans，发表于 *Evolution & Human Behavior* 24 (2003) 153—172，胡芸译。我们感谢 Martin Daly，Steve Frank 和 Margo Wilson 的评论给我们的帮助，并感谢桑塔费研究院和 John D. 及 Catherine T. MacArthur 基金的支持。

只能从表面上超越自私。在《道德系统的生物学》中，Alexander（1987，p.3）断言："只有把社会看作一个追求各自利益的个人集合时才能理解伦理、道德、人类行为和人类心理。"Ghiselin（1974，p.247）甚至声称："如果我们不感情用事，就会发现没有任何迹象表明纯粹的慈善行为会改善我们对社会的看法。所谓的合作成了机会主义和利用他人的结合体……抓住一个利他者，揭穿一个伪君子。"

然而，最近的实验研究揭示了人类行为的某些形式，这些人类行为包括很难用亲缘和互惠利他来解释的不相关个人间的交往。有一个我们称之为强互惠（Gintis，2000b；Herich et al.，2001）的行为，它的特征是与他人合作并花费个人成本去惩罚那些违反合作规范的人，甚至在预期这些成本得不到补偿或是在一个较迟时期才能得到补偿时也这么做。

在这篇文章中，我们拥有的证据支持强互惠行为是解释人类利他主义的关键。我们还会说明为什么在具有人类演化早期特点的条件下，较小数量的强互惠者会入侵自利者人群，以及为什么强互惠是一个演化稳定策略。尽管我们报道的大部分证据是建立在行为实验的基础上，但同样的行为也发生在日常生活中，比如公司工资的设定（Bewley，2000）、缴纳税款（Andreoni，Erard & Feinstein，1998），以及对地方环境公共物品保护中的合作（Acheson，1988；Ostrom，1990）。

我们不会为了支持强互惠的重要性而否认亲缘利他（Hamilton，1964）或互惠利他（Trivers，1971）的重要性。这两者毫无疑问是人类行为强有力的动机。然而，我们更相信物种的成功演化有赖于引导人们珍惜自由和平等的道德情操。而且，代议制的政府正是建立在强互惠和超越包括适存度及互惠利他的相关动机之上。

我们希望避免可能对我们的观点产生的三个普遍误解。首先，许多当代研究者反对我们对道金斯、Alexander以及其他"自私基因"学派的批评，他们声称他们的表述不能从字面上来理解。确切地说，作为表现型的"自私"行为应该理解为一种潜在的、受到达尔文进化力影响的基因结构。然而，这些作者们明明知道他们的表述很可能被错

误地理解，但仍然坚持使用夸张的陈述来表达一个“无懈可击”的命题。那么，这只能表明他们当时相信的这些陈述，现在看上去已经是不正确的了。

其次，我们经常被认为拒绝使用“基因中心”的方法为人类行为建模。事实上，我们的研究结果从来没有和研究基因与文化变迁的标准群体生物学方法相矛盾。一个为了他人牺牲自己利益的基因会消亡，除非这些基因会产生变异或是某些其他因素促进了这些基因的繁殖。在一个没有个体间社会交往结构的种群中，我们在实验中发现的及我们在模型中描述的行为者类型不会演化。然而，多层选择和基因—文化共生演化模型支持非亲缘间的合作（Bowles，Choi & Hopfensitz，出版中；Feldman，Cavalli-Sforza & Peck，1985；Gintis，2000，出版中a，b；Herich & Boyd，2001；Sober & Wilson，1998）。这些模型中的一部分会在下面讨论，而且这些模型不容易受到道金斯（1976），Maynard Smith（1976），Rogers（1990），Williams（1966）等人对群体选择理论的经典批评。

第三，我们经常被告知我们所描述的行为事实上可以用标准的个体选择、亲缘选择和互惠利他模型解释，这些模型适用于远古的、以我们这个物种演化初期为条件的自然和社会环境。在那个环境中，匿名、一次性的交往是非常罕见的。根据这个见解，那个时代环境下的强互惠行为是非适应性的。我们认为这是不可能的，并在本文第八部分提出我们的观点。

2 实验证据：劳动力市场的强互惠

在 Fehr、Gächter 和 Kirchsteiger（1997）的实验中，实验者把 141 个受试者（为了赚钱而同意参加实验的大学生）分成“雇主”集和一个更大一点的“雇员”集。博弈规则如下：如果一个雇主雇用一个提供

努力为 e 的雇员，雇员得到的工资为 w；雇主的利润 π 是努力 e 的一百倍减去他必须支付给雇员的工资 w （$\pi = 100e - w$），工资在 0 到 100 之间（$0 \leqslant w \leqslant 100$），努力在 0.1 到 1 之间（$0.1 \leqslant e \leqslant 1$）；雇员的支付 u 是他得到的工资减去“努力成本” $c(e)$，即 $u = w - c(e)$；努力成本 $c(e)$ 的大小由实验者预先规定，即当雇员提供的努力水平 $e = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ 和 1 时，实验者提供的努力成本表 $c(e) = 0, 1, 2, 4, 6, 8, 10, 12, 15$ 和 18。实验结束时，所有收入都以真实的现金支付给受试者。

行动顺序如下：雇主首先提供一个工资为 w、期望努力为 e^* 的合同，这个合同由最先同意这些条件的雇员达成，一个雇主最多可以和一个雇员订立（w, e^*）的合同；同意这些条件的雇员可以得到的工资为 w 并提供 e 的努力，但 e 并不一定要等于合同中的 e^*。事实上，如果雇员没有遵守诺言也不会受到惩罚，因此雇员可以选择任何水平的努力，$e \in [0, 1, 1]$。尽管受试者可能和不同的对手进行几次博弈，但每个雇主—雇员的交往是一次性事件，而且相互交往的同伴的身份不会被披露。

如果雇员是自利的，他们会选择零成本的努力水平 $e = 0.1$，不管雇主提供的工资为多少。由于知道雇主不会支付高于使雇员接受合同的最低工资 1（假设只允许整数的工资），雇员会接受这个合同并使 $e = 0.1$，此时的 $c(e) = 0$，那么雇员的支付 $u = 1$，雇主的利润 $\pi = 0.1 \times 100 - 1 = 9$。

然而，事实上，这个自利的结果很少在实验中出现。多次实验的结果显示，雇员的平均净收入 $u = 35$，而且雇主提供给雇员的工资越慷慨，雇员提供的努力水平就越高。实际上，如图 1 所示，雇主假设雇员有强互惠的倾向，提供较为慷慨的工资并得到更高水平的努力来作为提高他们自己和雇员收入的手段。Fehr，Kirchsteiger 和 Riedl（1993，1998）也发现了相似的结果。

图 1 还表明，虽然在任何一个工资率水平上大部分雇员都是强互惠

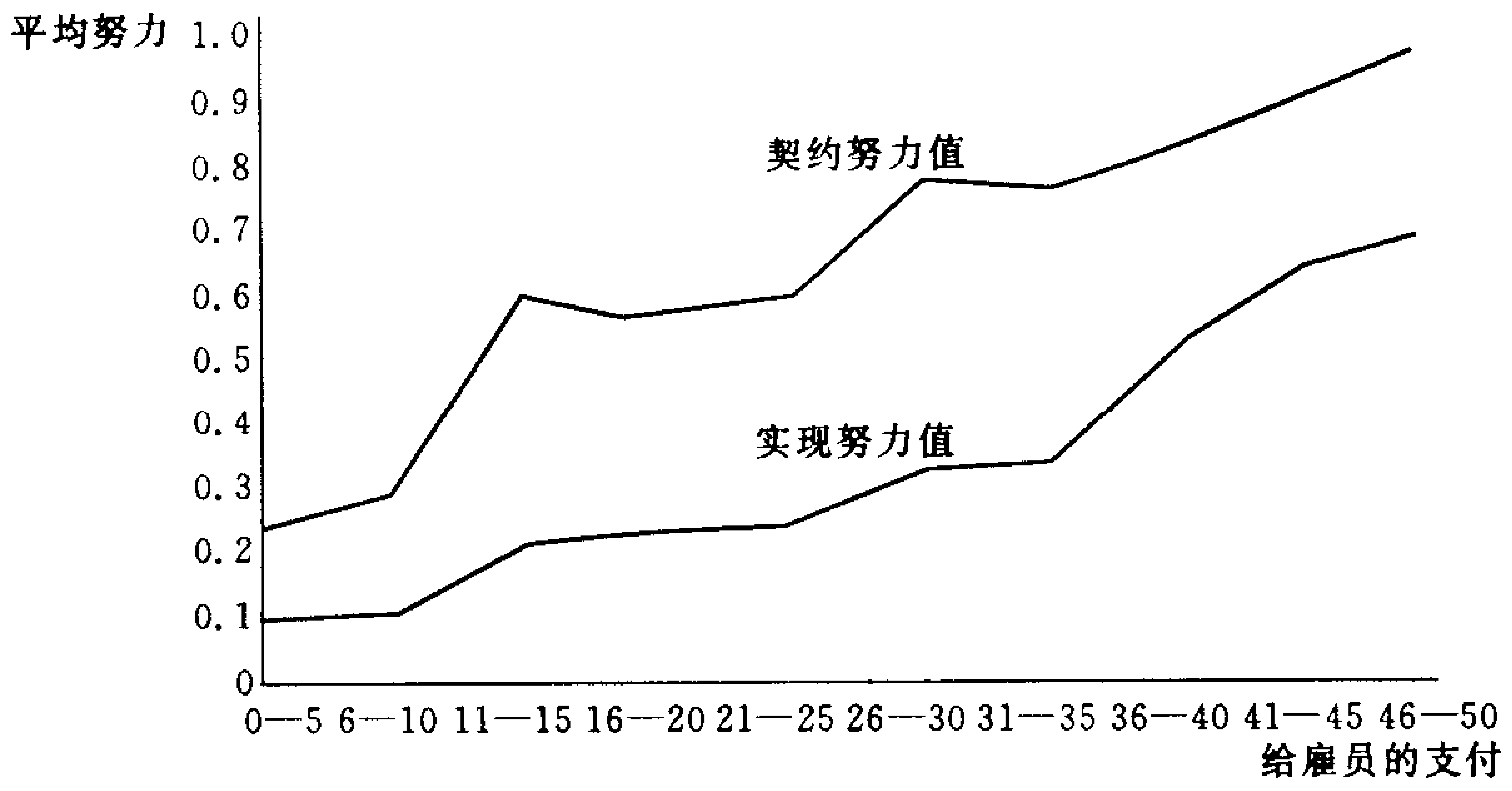

图1

者，但商定的努力水平和实际提供的努力水平之间还是存在很大的差距。这不是因为雇员集中有几个“坏家伙”，而是因为只有26%的雇员提供了他们承诺的努力水平！我们的结论是，即便是强互惠者，他们也会在某种程度上倾向于一种“折中”的道德水准，正如我们在日常经历中所预期的。

以上证据与雇主是纯粹自利的观念相容，因为雇主对雇员的善行在提高雇主自身利润上是有效的。在新一轮实验中，为了检验雇主是否也是强互惠者，实验者通过允许雇主对雇员的真实努力作出强互惠反应来拓展这个博弈。一个雇主可以通过支付1个单位的惩罚成本，使他的雇员的收入增加或下降2.5个单位。如果雇主是自利的，他当然不会采取任何行动，因为他们不会和同一个员工交往第二次。然而，实验结果显示，在68%的时间里，雇主会惩罚没有履行合同的雇员，在70%的时间里，雇主奖励了那些超额完成合同的雇员。事实上，雇主奖励了41%履行合同的雇员。而且雇员也预期到雇主的这种行为，如以下事实所述，即当他们的老板获得惩罚和奖励他们的权力时，雇员的努力水平得到了显著提高。未履行合同的人从83%下降到26%，超额完成合同的人从3%上升到38%。最后，即使把雇主对雇员惩罚的成本计算在内，允许雇主奖励和惩罚也能使所有受试者

的净收入增长40%。一些研究者（包括 Akerlof，1982；Blau，1964；Homans，1961）推测，这种行为是建立在普通、真实的社会生活基础上的。

我们从这个研究得出结论，扮演“雇员”角色的受试者遵循互惠的内化标准，甚至当他们知道以这种自身兴趣的方式采取行动不能获得物质回报时也这样做；而扮演“雇主”角色的受试者预期到这种行为，并且对这些行为进行相应的奖赏。最后，当允许“雇主”奖惩时，“雇主”会利用“奖好罚坏”的行为来内化规范；而“雇员”则会预期到这个行为，并相应地调整他们自己的努力水平。

3 实验证据：最后通牒博弈

在最后通牒博弈中，两个参与者在匿名条件下分配到一部分钱，比方说为10美元。一个参与者叫“提议者”，负责把这部分美元从1美元到10美元的任何一个份额分给另一个参与者，即“回应者”。提议者的提议只有一次，回应者在匿名条件下可以选择接受或拒绝对方提供的美元。如果回应者接受了提议，两人就按提议者的方案分配这部分钱。如果回应者拒绝，那么两个参与者都得不到钱。

由于博弈只进行一次，而且参与者并不知道对方的身份，一个自利的回应者应该接受任何数量为正的钱。知道这点之后，一个自利的提议者就会提供最小的数目1美元，而且这1美元会被对方接受。然而，在实验中，不但从来没有得到过这种自利的结果，并且从来没有接近过这个结果。事实上，正如这个实验的多次重复所表明的，在不同的环境和不同数量的钱的条件下，提议者通常提供给回应者的数目是非常可观的（标准的提供额是所有钱的50%），而回应者一般会拒绝低于总额30%的钱（Camerer & Thaler，1995；Güth & Tietz，1990；Roth，Prasnikar，Okuno-Fujiwara & Zamir，1991）。

在全球进行的最后通牒博弈，大部分参与者是大学生。我们从中发现了很多个体差异。比方说，在所有实验中有很大比例的受试者（一般大概为四分之一）以自利的方式采取行动。但是，在学生受试者中，不同国家间的平均行为惊人地一致。

为了增加实验受试者的文化和经济环境的多样性，Henrich et al.(2001) 在各种包括最后通牒博弈在内的博弈中，进行了一个大规模的跨文化行为研究。12 位富有经验的实地调查专家，在五大洲的 12 个国家，从 15 个有着极为不同的经济和文化环境的小规模社会中招募研究对象。我们的样本包括 3 个搜食社会（东非的 Hadza，巴布亚新几内亚的 Au 和 Gnau 以及印度尼西亚的 Lamalera），6 个刀耕火种的原始农业社会（Aché，Machiguenga，Quichua，南美的 Achuar，东非的 Tsimané 和 Orma），4 个游牧族群（Turguud，Mongols，中亚的 Kazakhs 以及东非的 Sangu），2 个定居的小规模农业社会（南美的 Mapuche 和非洲的 Zimbabwe 农民）。

我们可以概括出如下研究结论：

a．规范的自利行为模型在任何群体研究中都没有得到支持。以最后通牒博弈为例，在所有群体中，不管提议者还是回应者或是两者都表现出互惠的行为方式。

b．与原来的实验相比，跨文化研究发现了不同群体之间更多的行为差异性。最后通牒博弈中，学生受试者的平均出价在 43%到 48%之间，在我们的样本中，提议者出价的范围平均在 26%到 58%之间。在大学生中，始终按 50%出价的受试者，占统计样本的 15%到 50%。在某些群体里，即便是在出价非常低的情况下，拒绝的比例也非常少；而在另外一些群体里，拒绝的比例却非常大，包括经常拒绝超公正（hyperfair）的出价（例如超过 50%的出价）。与此相反，Machiguenga 最普遍的行为却是一毛不拔，平均出价是 22%。Aché 和 Tsimané 的分配有点类似美国人，但是拒绝率非常低。Orma 和 Huinca（在 Mapuche 中生活的非 Mapuche 智利人）的出价模式接近分配的中

心，但是比完全合作要低一点。

c. 不同群体在“市场整合”和“生产合作”中的不同程度，解释了群体间行为的部分差异：市场整合程度越高，合作的回报就越高，在博弈实验中合作和分享的水平就越高。这些族群被按照5个类别加以排列——“市场整合”（人们买卖和参加工作取得工资的频率），“生产合作”（集体生产还是个体生产），附加“匿名”者（匿名角色是否普遍以及如何处理），“隐私”（人们是否容易进行秘密行动）和“复杂性”（在家庭层面有多少决策是集权式的）。运用回归分析，只有最前面的两个特征（市场整合和生产合作）相关性非常显著，它们解释了最后通牒博弈中出价者行为差异的66%。

d. 个体层面的经济学和人口统计学的变量不能解释群体内部和群体间的行为差异。

e. 在实验中，合作和惩罚的性质与程度，通常跟这些群体日常生活中的经济形式是一致的。

在大量的案例中，博弈实验的开展和日常生活结构之间的相似性非常明显。被试主体在这种关系中对自我的地位有清醒的认识。下面是几个例子：

- Orma很快意识到公共物品博弈与harambee是一样的，harambee是当地家庭发起的建造公路或者学校的共同体决策。他们把这个实验称为“harambee博弈”，并且慷慨付出（平均有58%，其中25%给出了最大贡献）。

- 在Au和Gnau里，有些提议者给出高于一半的出价，而这些超公正的出价被拒绝了。这反映出美拉尼西亚通过馈赠礼物寻求社会地位的文化。在这些群体的日常生活中，大范围的赠予礼品是一种获取社会优势地位的要约，而拒绝礼品就是拒绝服从的意思。

- 在捕鲸族Lamalera中，最后通牒博弈中63%的提议者会对等出价，大多数人不会超过50%（平均出价是57%）。在现实生活中，个体捕鲸者之间通过合作捕获了一个大猎物时，会小心翼翼地取得事先约

定的份额，并分给共同体内的其他成员。

• 在Aché，79%的提议者会出价 40%或者 50%，并且有 16%的人出价高于 50%，这些出价不会被拒绝。在日常生活中，Aché 人分享肉类，平等地在各个家庭之间分配而不考虑到底是谁打到了这个猎物。

• Hadza 和Aché 不一样，在最后通牒博弈中常常出价比较低，并且拒绝率很高。这反映出只在小范围的狩猎者内部分享肉类的趋势，群体内的冲突很激烈，而且猎人经常会把自己的猎物藏匿起来。

• Machiguenga 和Tsimané 在最后通牒博弈中出价都很低，而且事实上都没有被拒绝。在家庭以外，这些群体很少进行合作、交易和分享。从人类学意义上看，他们也都表现得很少敬畏社会制裁，很少关心“公共评价”。

• Mapuche 人的社会关系的特征是相互猜疑、嫉妒并且害怕被嫉妒。这种模式和 Mapuche 人在最后通牒博弈后的会面是一致的。Mapuche 提议者很少要求出价考虑公平，但是却害怕被拒绝。甚至提议者提出超公平的出价仍然声称害怕极少数恶意的回应者，因为那些人甚至会拒绝对等出价。

4 实验证据：公共品博弈

有一系列的论文分析公共品博弈，如社会心理学家 Toshio Yamagishi (1986, 1988)，政治科学家 Elinor Ostrom 及其合作者（Ostrom, Walker and Gardner, 1992），经济学家 Ernst Fehr 及其合作者（Gächter and Fehr, 1999; Fehr and Gächter, 2000, 2002）的论文。这些研究者无一例外地发现，相对于假设的、能被预期的标准经济学模型里的自利人，当受试者搭上自己的费用去惩罚搭便车者的时候，群体展示出了更高的合作率。

一个典型的公共品博弈由几轮构成，比如说 10 轮。受试者事先

被告知博弈总共进行的次数，也被告知有关博弈的其他各个方面的信息，博弈结束后受试者赢了多少就能获得相应的真实货币。在每一轮，每一个受试者都和其他几位——比如三个——组成一个严格匿名的群体。每个受试者会有一个确定的“点数”，比如 20，在实验的最后阶段可以兑换成真正的钱。每一个受试者可以把他的一部分点数放在一个“公共账户”里，其他放在自己的“私人账户”里。然后，实验者告诉受试者公共账户里有多少点数，每个私人账户最后都能得到公共账户里总点数的 40%。如果一个受试者把其所有的 20 个点数都捐给了公共账户，群体中的每一个成员每轮都能拿到 8 个点数。实际上，一个参与者把所有的都捐给公共账户，他所损失的是 20 个点数；但这一轮结束后，不管私人账户里有多少，其他三个成员总能够获得总数为 24 （8 × 3） 个点数。

一个自利的参与者不捐助公共账户，可是实际上只有少数受试者符合自利模型。受试者起初平均贡献他们点数的一半给公共账户，贡献水平随着博弈轮次的运行而减少，直到最后一轮绝大多数参与者都以自利方式行为（Dawes and Thaler， 1988； Ledyard， 1995）。在 12 个公共品博弈实验中，Fehr 和 Schmidt （1999）发现在前面几轮中，均值和中值的捐助贡献水平在 40%到 60%之间。但是在最后阶段，73%（总数是 1 042）的个体没有任何贡献，而剩下的参与者的贡献也接近 0。这个结果和自利人的模型是一致的，自利人模型预测每一轮所有人的贡献都是零。虽然也可以用互惠利他模型来预测，但实验的结果是互惠的减少。实际上这并没有解释适度的合作水平，反而解释了在公共品博弈中合作水平的恶化。

实验后对受试者合作出价减少的解释报告是这样的，那些合作的受试者对比他贡献少的人表现出愤怒之情，然后利用惟一的手段——减少他们自己的贡献——来报复那些低贡献的搭便车者（Andreoni， 1995）。

实验证据支持了这个解释。当所有受试者都被允许惩罚无贡献者时，他们会付出成本来坚持这样做（Dawes， Orbell and Van de

Kragt, 1986; Sato, 1987; Yamagishi, 1988a, 1988b, 1992)。举例而言，在Ostrom等人(1992)的实验中，受试者在一个25轮的公共品博弈中相互作用，通过支付一定的“费用”，可以对其他人“罚款”以增加他人的成本。罚款成本的使用，自然会增加整个群体的利益。但这个博弈里惟一的纳什均衡不依靠这个可疑的威胁，因为没有参与者愿意支付这个“费用”，于是没有人会因为背叛而受到惩罚，这样所有人都会背叛而不为公共领域贡献任何东西。然而研究者发现，惩罚行为非常显著。

这些研究允许个人使用策略行为，因为对背叛者的高成本惩罚会在将来增进合作，对执行惩罚者会有一个正的净回报。Fehr和Gächter (2000)设置了一个没有策略性惩罚的实验情景。他们使用6轮和10轮的公共品博弈，群体的规模是4个人，只允许最后一轮才能施行高成本的惩罚。他们采用三种不同的指派成员的方法，有足够多的受试者在10到18个群体之间同时进行实验。在所谓“伙伴待遇”的情况下，全部10轮实验中，4个受试者始终都在同一个群体里。在所谓“陌生人待遇”的情况下，每一轮结束后受试者都会重新随机组合。最后，在所谓“完全陌生人待遇”的情况下，受试者会随机组合并且保证他们不会再遇到同一个受试者。每一个受试者在开始实验前大约平均可以得到35美元。

Fehr和Gächter (2000)进行了10轮有惩罚的实验和10轮没有惩罚的实验。他们的结论可以用图2来表示。

我们看到，允许高成本的惩罚时，合作没有恶化。在伙伴待遇组中，尽管严格匿名，但合作却增强并且几乎到了完全合作的境况，甚至在最后一轮也是这样。当不允许惩罚时，在早先的公共品博弈中同样的受试者经历了合作的恶化。在伙伴待遇和两种陌生人待遇情况下的合作率的对比是值得注意的，因为惩罚的强度在所有待遇中大致相同。这显示出惩罚在伙伴待遇情况下更有效，因为在这种待遇情况下，被惩罚者确信，一旦当他在前面几轮被惩罚了，这个群体里就有了执行惩罚

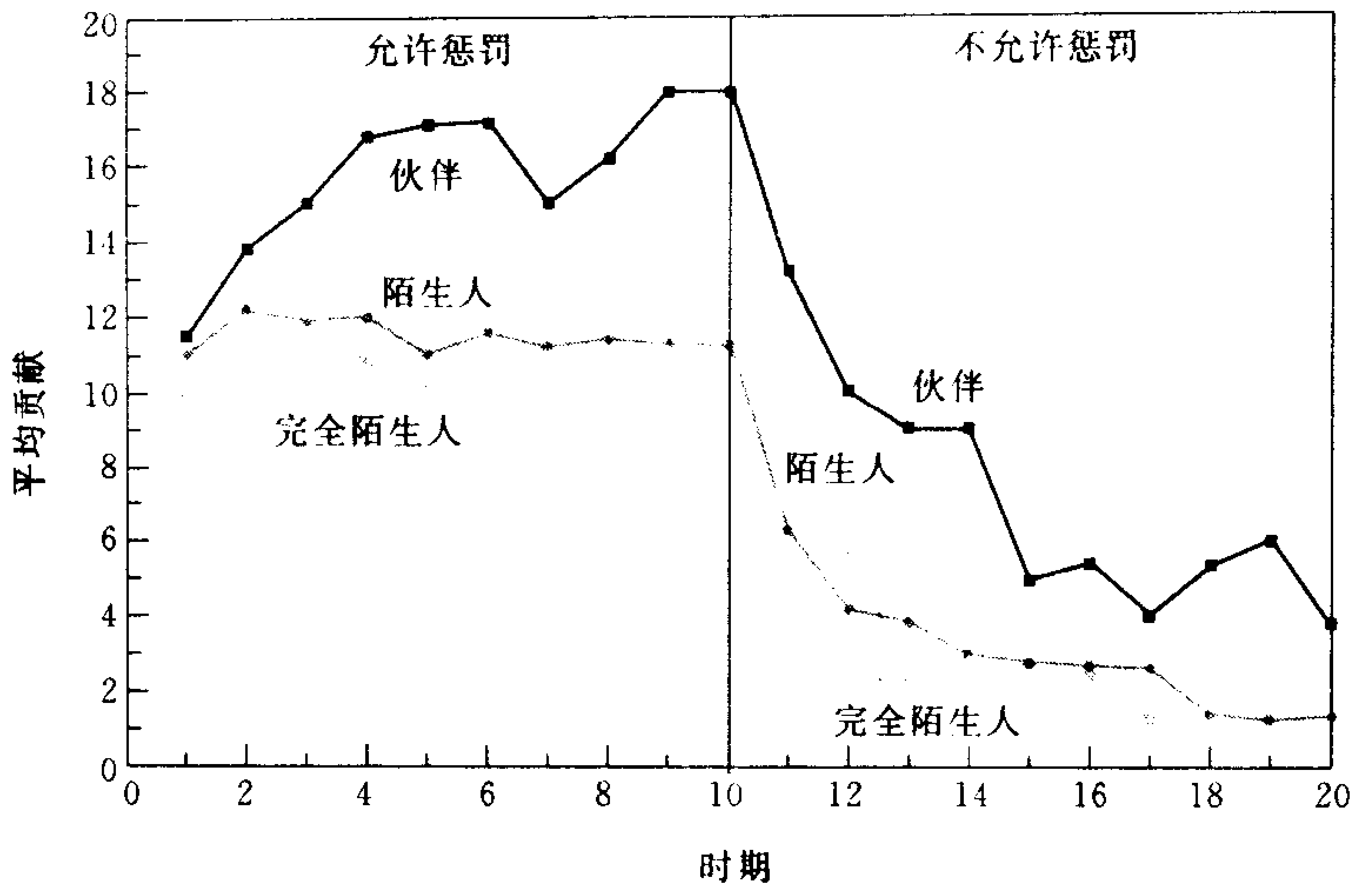

图 2

者。这就更加证实了强互惠的趋社会性对合作产生了影响，被研究的群体变得更加团结并能使合作维持得更久。

5 实验证据：意图还是结果？

上面展示的实验中遗漏的一个关键要素是对贡献和惩罚关系的界定。强互惠的解释表明高贡献者将是高惩罚者，而被惩罚的人是贡献低于平均水平者。这个推测是由 Fehr 和 Gächter（2002）证实的，75%的惩罚行为是由贡献高于平均水平者实施的，预测一个参与者对另一个参与者实施多大惩罚的最重要变量是惩罚者自己的贡献量和被惩罚者贡献量之间的差异。

理解上述证据的另一个主要问题是：互惠者是对公平或不公平的意图作出反应还是对公平或不公平的结果作出反应？毫无疑问，强互惠模型对意图的重视超过对结果的重视。为了回答这个问题，Falk, Fehr和 Fischbacher（2002）进行了两个版本的“偷袭博弈”——一个有

意图方式（I），但是在这里不能从他的行为中推断出这个参与者的意图，以及一个无意图方式（NI），在这里可以从一个参与者的行为中推断他的意图。他们为消极和积极互惠行为提供了意图和行为相关的明确证据。

偷袭博弈包括两个阶段。在博弈一开始，两个参与者都得到12点代币。在第一阶段，参与者A选择一个行为 $a \in \{-6, -5, \cdots, 5, 6\}$。如果A选择 $a > 0$，他给参与者B的代币为 a，但是如果他选择 $a < 0$，那么他从参与者B中拿走的代币为 $|a|$。在 $a \geqslant 0$ 的情况下，实验者把三倍的 a 给B，那么B就得到 $3a$。当B发现 a 的时候，他可以选择一个行动 $b \in \{-6, -5, \cdots, 17, 18\}$。如果 $b \geqslant 0$，B把 b 数量的代币给A。如果 $b < 0$，B失去 $|b|$ 而且A失去 $|3b|$。由于A可以给予或拿走，而B可以奖励或惩罚，那么这个博弈既允许积极互惠行为也允许消极互惠行为。每个受试者只能参加一次博弈。

如果参与者B是自利的，那么他们都会选择 $b = 0$，既不奖励也不惩罚他们的同伴A，因为这样的博弈只进行一次。知道这点之后，如果参与者A是自利的，那么他们都会选择 $a = -6$，这时他们最大化自己的支付。在I条件中，允许参与者A选择 a；但是在NI条件下，参与者A的选择由一副配对的骰子决定。如果参与者不是自利的而且只关注公平的结果而不是意图，那么参与者B在I和NI条件下的行为是相同的。而且，如果参与者A认为他们的同伴B只关注结果，那么两种情况下参与者A的行为也不会改变。如果参与者B只关注同伴A的意图，那么在NI条件下B将永远不会奖励或惩罚，但是在I条件下将奖励选择较高 a 的同伴以及惩罚选择较低 a 的同伴。

实验者的主要结果是参与者B在I条件下和在NI条件下的行为有很大差异，这表明对公平意图的关注具有行为上的重要意义。事实上，在I条件下参与者B因从A处获取而给予A的奖励要高于在NI条件下B给予A的奖励（显著性为 $P < 0.01$），而且在I条件下A因从B处获取而受到的惩罚要高于在NI条件下的惩罚（显著性为 $P < 0.01$）。

回到个人的行为模式，在I条件下，没有一个行为者的行为是纯粹自私的，然而，在NI条件下，30个人的行为是完全自私的。在I条件下，76个受试者奖励或制裁了他们的同伴，而在NI条件下，只有39个受试者奖励或惩罚了他们的同伴。我们得出的结论是，大部分的行为者受他们同伴意图的驱动，但有显著比例的受试者要么只关注结果要么同时也关注同伴的意图。

6 强互惠的稳定演化

金迪斯（2000b）在论文中提出了一个解析模型，该模型显示，在可行的条件下，强互惠可以通过群体选择在互惠利他主义中产生。这篇论文中，合作被模拟为在一般条件下，行为者在对将来能从其他成员那里得到回报有充分认识的情况下的 n-人公共品重复博弈。合作通过放逐威胁得以持续。但是当族群面临灭绝或者解散的威胁时，比如说遇到战争、瘟疫或者饥荒，为了生存，合作将变得更为必需。然而，当族群受到威胁时，一个人对族群的贡献在将来得到回报的可能性会急剧下降。随着族群解散的可能性上升，合作的动机也将不复存在。因而，恰恰当一个族群最需要趋社会行为时，基于互惠利他主义的合作将面临崩溃。

少量不考虑未来回报而对背叛者施以惩罚的强互惠者能够显著提高人类族群的生存机会。而且，人类是惟一具有以低成本对受罚者施加严厉处罚能力（Bingham，1999）的群居物种，这是人类具有高超的制造工具和狩猎能力（Darlington，1975；Fifer，1987；Goodall，1964；Isaac，1987；Plooij，1978）的结果。确实，同非人类灵长类作严格比较，在这些条件下，强互惠者能够侵入一个利己类型的人群并保持均衡。这是因为，即使人群中只有一小部分的强互惠者，至少他们有时可以组成一个较大比例的族群使得合作在艰难的时候得以保持。

然后这样的族群将会胜出其他的自利族群，从而使强互惠者的比例得到增加。这个现象会继续，直至达到强互惠者的均衡比例。

尽管可以解析地得到上面的结果，却不容易对强互惠者均衡比例作数学描述。然而，计算机模拟非常能说明问题。比如说，假设在好年景时一个族群在一个阶段的幸存机会是95%，在年景不好的时候（10个阶段中的发生一次），族群在这个阶段的幸存机会是25%。图3中较低的曲线表明，当每个族群有40个成员时，惩罚成本 c（r）的变化与强互惠者均衡比例 $\bar{f}^*$ 的关系。较高的那条曲线表明了当每个族群有8[1]个成员时的相同关系。如果族群是由几个“家庭”组成的，同时强互惠的特点在家庭中具有高度的传递性，那么一个很小比例的强互惠者就能确保合作，而且惩罚的成本很低，强互惠者的均衡比例很大。

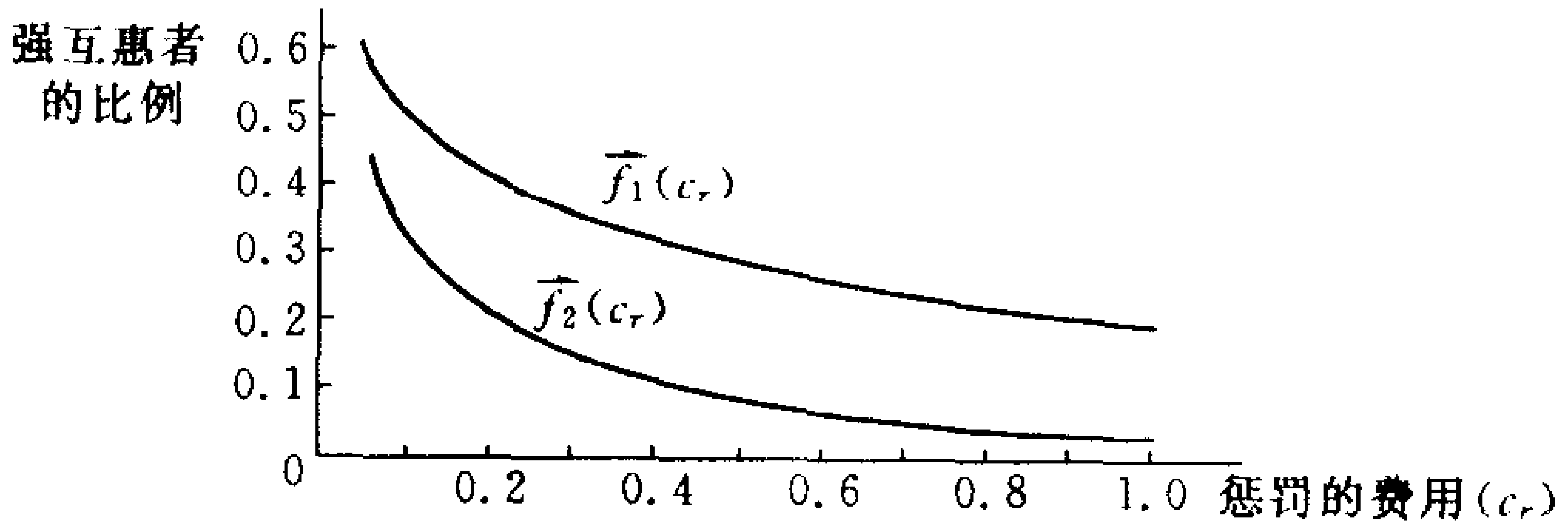

图3

这个模型强调了强互惠的一个关键的适应性特征——它独立于未来交往的可能性——互惠利他只有预设未来交往的可能性较高时才能解释一般的合作。然而，互惠利他在大的族群中是缺乏效率的（Boyd & Richerson，1998；Choi，2002；Joshi，1987；Taylor，1976）。这是因为，当一个人为报复族群中的一个背叛成员而不进行合作时，事实上他对所有的成员包括背叛者和合作者都施加了惩罚。n-人公共品博弈惟一的稳定演化策略是：只有当其他所有人合作时才合作，而在其中一个人背叛时就背叛。对任何一个这样支付单调的动态系统来说，均衡的凝聚力在族群规模上升时变得非常微小，因此该族群由较大数目的、有条件的合作者组成的可能性很小，因为这样一个结果随时可能被异质的参

与、对于其他参与者而言的不完美信息或任何随机事件所打断。所以，如果族群的规模很大，这种均衡不可能在某个合理的历史时刻出现。而且，只要有一个成员背离，这种惟一的“边缘”均衡也将崩溃。

为了使强互惠的演化更具现实性，Henrich 和 Boyd（2001）提出了一个模型，在这个模型中合作和惩罚的规范通过支付倾斜传递（模仿成功者）和顺应传递（模仿多数行为）来获得。他们表明，如果两个阶段都允许惩罚，那么通过稳定的惩罚，可以在一个任意小数量的顺应传递者之间保持牢固的合作行为。接着，他们解释了一旦合作在一个族群中趋于稳定，它是如何通过文化的群体选择功能在一个多个群体的人群中传播的。一旦合作盛行，他们进一步指出有利于合作和惩罚的趋社会基因如何入侵文化的群体选择。比如，因为这样的基因减少了高成本惩罚者所冒的风险。

上述分析揭示了利他合作和利他惩罚之间很强的不对称性，Boyd, Gintis, Bowels 和 Richerson（2002）进行了进一步的研究。他们的研究表明，利他惩罚能够在一个更大的族群中导致合作，因为利他合作者相对于背叛者的支付劣势是独立于人群中背叛者的比例的；而当背叛者变得稀少时，用于利他惩罚所造成的成本劣势将趋于下降，因而，这个时候利他惩罚就会很普遍，不利于他们的选择压力也较弱。当背叛者很少时，惩罚者只面临很小的成本劣势，这一事实意味着弱的族群内部的演化力量，如顺应传递，能使惩罚变得稳定并允许维持合作。计算机模拟显示，当族群不能维持利他合作时，族群间的选择将导致利他惩罚的演化。

7 制度与行为的共生演化

如果群体选择是对个体合作行为成功演化的部分解释，那么可以认为增强群体选择压力的特征（比如，相对较小的族群规模，有限的移

民，频繁的群间竞争）是与合作行为共同演化的，这种族群层面的特征与个体行为可能存在着协同效果。有这样的例子，合作是部分基于一种只有人类才具有的特殊能力，它可以构建某种制度环境，限制群内竞争、减少群内非典型变异，同时提升群间竞争的重要性，并允许个别贵价、但有利于群内的行为通过群间选择过程同那些支持性的制度环境共生演化。

这种对群内竞争进行压制可能会严重影响进化动力的观点可以在群居昆虫或其他物种中得到广泛证实。Boehm（1982）和 Eibl-Eibesfeldt（1982）首先将这种推理运用到人类社会，探究文化传递在减少群内显型变异中的作用。这些惯例比如均等制度，在非亲属间分享资源，以及减少群内福利或物质福利的差异。这些惯例可以减少在福利或物质福利缺乏时，群内成员所获资源过于严重的不平等。因此，尽管对群内成员表示慷慨的优秀猎人可能比其他猎人拥有较好的福利和营养（作为消费结果比较），但这并不表明缺乏均等，除非这些惯例会导致不太成功的猎人的福利和营养恶化（这似乎不太可能）。通过减少成功个体方面的群内差异，这些惯例可能会削弱族群内基因或文化选择上的操作，这些操作反对对族群有利而对个人来说要支付代价的举措，因而使族群得以在群间竞争时获得更大的利益。族群层次的制度因此创造了某种新的环境，这种环境能够展现生物进化和文化变迁过程的方向和速度。许多增强选择压力从而降低族群灭绝可能性的事实，可以解释减少群内表型变异的社会制度的成功演化。

我们沿着这些新的特征建立一个演化动力学模型。这个模型通过文化、基因传递个人行为以及通过文化传递族群层次的制度特征，使群间竞争对族群层次的选择起决定作用（Bowles，2001；Bowles，Choi & Hopfensitz，2002）。我们可以看到，群间竞争可以成功地解释下面两个进化：(a) 人类社会的利他形式从亲缘转向非亲缘；(b) 族群层次上的制度结构，比如在人类漫长的历史进程中重复出现并逐步扩散的资源分享。如果对群外成员施加惩罚的成本足够高，而族群层次的制

度能够限制这些行为，因此削弱了反对这些行为的群内选择，那么有利于族群内部利益的行为就可能得到进化。

我们的模拟显示，如果在族群层次上贯彻资源共享或者族群成员非随机配对的制度能够进化，那么有利于族群的个体特征就能与这些制度共生演化，即便后者会给采纳它们的族群带来高额成本。这些结果在下面的条件中保持不变：族群中个体合作行为和社会制度在一开始是缺失的。但是，在缺乏族群制度时，只有当群间的冲突非常频繁、族群规模较小且移民比例较低时，有利于族群的特征才能进化。因此，合作行为成功进化的相关环境，是存在于9万年前解剖学意义上的近代人类所特有的建构社会制度能力的结果（Boyd & Richerson，出版中）。

8 对强互惠的其他解释

我们认为很多实验数据支持我们关于强互惠行为的假设，而且，根据当前和古代的证据，强互惠行为在基因与文化共生演化过程中是具有适应性的。第一个论断遭到 Price，Cosmides 和 Tobby（2002）的批评，他们对数据作出了另一种解释。第二个论断也遭到许多人的批评，他们认为博弈实验中利他合作和利他惩罚是不利适应的反应，因为受试者所面临的实验场景不论在人类演化史或是当前的日常生活中都没有，人类不会对这些场景作出适应性的基因或文化的反应。最强烈的抨击是认为我们所描述的利他行为对理解自然状态下的人类行为没有重要意义。我们依次来解决这两个批评。

8.1 强互惠与适存度赤字的减少

Price 等（2002）提出了一个他们认为能够解释实验博弈中的行为的自利模型。尤其是，他们断言（p. 221）：“惩罚搭便车者的动机是为了防止出现搭便车者的适存度优势超过贡献者的适存度优势。”

我们将指出（a）Price 等的模型和现存的实验数据不相容；（b）Price 等的模型不是演化稳定的。

Price 等的模型断言，惩罚性行为是对支付差异而不是对违反互惠规范的反应。实验者检验了惩罚意愿是否是对支付差异的反应，在适当控制的场景下，结果几乎是清一色的否定。要了解这个问题的全面讨论，参看 Falk，Fehr 和 Fischbacher（2001）的文献。

比如说，我们把最后通牒博弈中提议者在 10 美元中分给回应者的钱限定在 2 美元或 8 美元，而且回应者也知道这个限定。接着，尽管在未限定的情况下回应者会拒绝 2 美元，但在限定条件下几乎所有回应者都接受了提议者提供的 2 美元。这和回应者采取行为是为了减少适存度差异的假设不相容，因为在非限定和限定情况下这个目的都能实现。但是，实验中出现的行为与回应者采取行动是为了惩罚不慷慨的提议者的假设一致。因为在限定的情况下，提议者无法使自己变得慷慨，除非采取极端的或是不合理的慷慨——慷慨到他自己只保留 2 美元。

在另外一个实验中（Blount，1995），最后通牒博弈中提供的钱由计算机而不是提议者给出，并且回应者知道这个事实。在这个情况下，甚至数额很低的钱也很少被拒绝。这和强互惠假设相容，因为回应者没有动机去惩罚不对低支付负责的提议者。然而，根据 Price 等的模型预测，不管由计算机还是提议者来分钱，低支付都会被拒绝，因为这两者都给提议者带来了相同的适存度优势。

其实，更一般的情况是，在标准的最后通牒博弈中（参见第三部分），一个接受 50% 以下任何份额的回应者都会产生一个相对的适存度损失，然而几乎所有回应者都会接受 40% 的份额。相似的，在雇员—雇主博弈中（参见第二部分），雇主自愿把资源转移给雇员。

回到演化稳定的问题上来，Price 等的模型是建立在选择更喜欢惩罚性情感的假设上的，因为这样的情感会减少背叛者相对的适存度优势。正如作者所声称的，选择取决于相对而不是绝对的适存度，因此

任何减少承受者适存度的“恶意”行为，在理论上都会增加惩罚的频率，只要这个行为使其他人的适存度比自己减少得更多。但是——这是十分重要的——衡量相对适存度不仅包括他们惩罚的目标，还包括那些合作充分从而免受惩罚且又不承担惩罚成本的其他相关人。这些“远离麻烦”的行为者将会比惩罚者（或他们的目标）具有更高的适存度，从而会以其他行为者为代价来扩展他们在人群中的份额。

用人类学的术语说，其他相关人与自己是属于一个同类群的个人（在同一个演化人群中生活和繁殖的个人），而不是特殊个人的社会群体（在同一个演化人群中专门交往的个人）。由于这个分类，如果一个人群足够小，从而这个人群中的社会群体和同类群的大小是一致的，选择将更青睐恶意的行为（Hamilton, 1970），即使在趋社会的物种（Foster, Wenseleers, & Ratnieks, 2001）中也是如此。但是人类搜食社会的证据表明，同类群的规模比社会群体大，因此恶意行为不会被选择。

解析模型表明，如果社会群体的成员在完全零和的条件下竞争，那么恶意行为会演化，而且一个人适存度的增加必然意味着这个社会群体中其他成员的再生产产出会因补偿而减少（Boyd, 1982）。但我们知道，在人类学意义上的狩猎—采集社会生产力水平下，上面的情况是不太可能发生的。在这样的社会中，资源没有本土化，群体是流动的，而且部落间的个人迁徙频繁，因此在很大程度上违反了完全零和的条件。

一个社会群体的子集，为了争夺有限的资源，在一个近乎零和的条件下竞争是可能的。比方说，男人可能为了女人而竞争，或者为可以转化成相对适存度的优势地位而竞争。在这种情况下，选择可能会更青睐减少竞争对手适存度的恶意行为。然而，在这种情况下，“认知回路”马上就会被构建起来，从而使这个人只致力于减少与他有特殊零和竞争关系的、特定的个人的适存度。举个例子，男人应该关心与他年龄相仿的男人，那些人的地位、等级与他相当，但不关心比他

大的或比他小的男人以及女人。这个情况不适用于平等分享合作成果的公共品博弈。

8.2 强互惠与非适应性

一些行为学的科学家认为我们在这篇文章中描述的行为是不利适应性的，而且和真实的社会互动不相关。他们提出，人类的大脑不是一个一般的目的性信息处理器，而是一套用于解决在我们这个物种演化历史中面临的特殊问题的互动模块。由于实验博弈匿名、不重复互动的特点，我们无法预期受试者在实验博弈中会按最大化适存度采取行动。因此我们必须在受试者的实验环境中引进一些与演化历史相似的条件，像非匿名、重复互动之类的信息，然后才能确定这个被重新解释的环境与最大化适存度之间的关系。

这个批评如果是正确的，也不会减少强互惠行为在当代社会中的重要性。在某种程度上，现代生活经常性地使个人面临匿名的、非重复的互动，这是具有先进贸易、通信和交通技术的现代社会的特征。因此，如果强互惠行为曾经是非适应性的，但它仍然是解释现代人类合作的一个重要因素。

但我们不相信这个批评是正确的。事实上，人类完全有能力分辨其他人在将来与自己交往次数的多少。这样，他们预期将来和他人有较多的交往时所采取的合作，会比他们预期将来和他人有较少交往机会时的合作更加充分（Gächter & Falk，2002；Keser & van Winden，2000）。

在我们祖先的生活环境中，具有精致协调的行为系统的人类很可能获得某种进化的优势，不管他们面临的是亲缘还是非亲缘、重复还是一次性的互动，也不管这么做是否能增加个人的声誉；这个优势产生的原因更可能是因为人类面临许多互动，而在这种互动中在未来继续互动的概率非常低，以至于背叛是值得的（Gintis，2000b；Manson & Wrangham，1991）。

9 结 论

为了理解人类趋社会行为的主要脉络，还需进行更多的实验和理论工作。我们猜想，在过去几年已经完成的许多研究的基础上，所获得的新知识会给我们一个趋社会的全貌（它的反面是反社会），这个趋社会的理论从根本上和自利的经济学模型与自涉的互惠利他生物学模型不相容。

当代行为学理论是几个重要理论贡献的遗产（包括 Cosmides & Tooby, 1992; Dawkins, 1989; Hamilton, 1964; Maynard Smith, 1982; Trivers, 1971; Williams, 1966; Wilson, 1975），他们所有人都假设可以使用自利模型来解释非亲缘个体之间的关系。因此行为学理论最成功的研究是在家庭、亲缘关系和性关系领域，而在处理具有更复杂互动特征的社会群体行为方面，没有什么大的说服力。这一点都不足为怪。为了说明这些情况，我们认为应该更注意（a）社会情感的起源和本质（包括愧疚、羞耻、同情、种族身份和种族仇恨）；（b）人类社会历史的基因与文化的共生演化；（c）人类演化中的族群结构和族群冲突；以及（d）把社会生物学的见解融入主流的社会科学。

注释：

[1] 原文如此，疑为 80。——译者注

参考文献：

Acheson, J. (1988). *The Lobster Gangs of Maine*. Hanover, NH: New England Universities Press.

Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *Quarterly Journal of Economics*, 97, 543—569.

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York: Aldine.

Andreoni, J. (1995). Cooperation in Public Goods Experiments: Kindness or Confusion. *American Economic Review*, 85, 891—904.

Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax Compliance. *Journal of Economic Literature*, 36, 818—860.

Bewley, T. F. (2000). *Why Wages don't Fall During a Recession*. Cambridge: Harvard University Press.

Bingham, P. M. (1999). Human Uniqueness: a General Theory. *Quarterly Review of Biology*, 74, 133—169.

Blau, P. (1964). *Exchange and Power in Social Life*. New York: Wiley.

Blount, S. (1995). When Social Outcomes aren't Fair. The Effect of Causal Attributions on Preferences. *Organizational Behavior & Human Decision Processes*, 63, 131—144.

Boehm, C. (1982). The Evolutionary Development of Morality as an Effect of Dominance Behavior and Conflict Interference. *Journal of Social and Biological Strue tures*, 5, 413—421.

Bowles, S. (2001). Individual Interactions, Group Conflicts and the Evolution of Preferences. In: S. N. Durlauf, & H. P. Young (Eds.), *Social Dynamics* (pp. 155—190). Cambridge, MA: MIT Press.

Bowles, S., Choi, J., & Hopfensitz, A. (in press). *The Co-evolution of Individual Behaviors and Social Institutions*. Santa Fe, NM: Santa Fe Institute.

Bowles, S., & Gintis, H. (2002). Homo Reciprocans. *Nature*, 415, 125—128.

Boyd, R. (1982). Density Dependent Mortality and the Evolution of Social Behavior. *Animal Behavior*, 30, 972—982.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2002). *Altruistic Punishment in Large Groups Evolves by Interdemic Group Selection*. Unpublished data.

Boyd, R., & Richerson, P. (in press). *The Nature of Cultures*. Department of Anthropology, UCLA, Los Angeles, CA.

Boyd, R., & Richerson, P. J. (1988). The Evolution of Reciprocity in Sizable Groups. *Journal of Theoretical Biology*, 132, 337—356.

Camerer, C., & Thaler, R. (1995). Ultimatums, Dictators, and Manners. *Journal of Economic Perspectives*, 9, 209—219.

Choi, J.-K. (2002). *Three Essays on the Evolution of Cooperation*. Amherst, MA: University of Massachusetts.

Cosmides, L., & Tooby, J. (1992). The Psychological Foundations of Culture. In: J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 19—136). New York: Oxford University Press.

Darlington, P. J. (1975). Group Selection, Altruism, Reinforcement and Throwing in Human Evolution. *Proceedings of The National Academy of Sciences, U. S. A.*, 72, 3748—3752.

Dawes, R. M., Orbell, J. M., & Van de Kragt, J. C. (1986). Organizing Groups for Collective Action. *American Political Science Review*, 80, 1171—1185.

Dawes, R. M., & Thaler, R. (1988). Cooperation. *Journal of Economic Perspectives*, 2, 187—197.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, R. (1989). *The Selfish Gene* (2nd ed.). Oxford: Oxford University Press.

Eibl-Eibesfeldt, I. (1982). Warfare, Man's Indoctrinability and Group Selection. *Journal of Comparative Ethnology*, 60, 177—198.

Falk, A., Fehr, E., & Fischbacher, U. (2001). *Driving Forces of Informal Sanctions*. Working Paper No. 59, Institute for Empirical Research in Economics.

Falk, A., Fehr, E, & Fischbacher, U. (2002). *Testing Theories of Fairness and Reciprocity —Intentions Matter*. Zürich: University of Zürich.

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment. *American*

Economic Review, 90, 980—994.

Fehr, E. , & Gächter, S. (2002) . Altruistic Punishment in Humans. *Nature*, 415, 137—140.

Fehr, E. , Gächter, S. , & Kirchsteiger, G. (1997) . Reciprocity as a Contract Enforcement Device: Experimental Evidence. *Econometrica*, 65, 833—860.

Fehr, E. , Kirchsteiger, G. , & Riedl, A. (1993) . Does Fairness Prevent Market Clearing? *Quarterly Journal of Economics*, 108, 437—459.

Fehr, E. , Kirchsteiger, G. , & Riedl, A. (1998) . Gift Exchange and Reciprocity in Competitive Experimental Markets. *European Economic Review*, 42, 1—34.

Fehr, E. , & Schmidt, K. M. (1999) . A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114, 817—868.

Feldman, M. W. , Cavalli-Sforza, L. L. & Peck, J. R. (1985) . Gene — Culture Coevolution: Models for the Evolution of Altruism with Cultural Transmission. *Proceedings of the National Academy of Sciences*, *U. S. A.* , 82, 5814—5818.

Fifer, F. C. (1987) . The Adoption of Bipedalism by the Hominids: a New Hypothesis. *Human Evolution*, 2, 135—147.

Foster, K. R. , Wenseleers, T. & Ratnieks, F. I. W. (2001) . Spite: Hamilton's Unproven Theory. *Annales Zoologici Fennici*, 38, 229—238.

Fudenberg, D. & Maskin, F. (1986) . The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54, 533—554.

Gächter, S. & Falk, A. (2002) . Reputation or Reciprocity? Consequences for Labour Relations. *Scandinavian Journal of Economics*, 104, 1—25.

Gächter, S. & Fehr, E. (1999) . Collective Action as a Social Exchange. *Journal of Economic Behavior and Organization*, 39, 341—369.

Ghiselin, M. T. (1974) . *The Economy of Nature and the Evolution of Sex*. Berkeley, CA: University of California Press.

Gintis, H. (2000a) . *Game Theory Evolving*. Princeton, NJ: Princeton University Press.

Gintis, H. (2000b) . Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology*, 206, 169—179.

Gintis, H. (in press-a) . The Hitchhiker's Guide to Altruism: Genes and Culture, and the Internalization of Norms. *Journal of Theoretical Biology*.

Gintis, H. (in press-b) . The Puzzle of Human Prosociality. *Rationality and Society*, 15.

Goodall, J. (1964) . Tool-Using and Aimed Throwing in a Community of Free-Living Chimpanzees. *Nature*, 201, 1264—1266.

Güth, W. & Tietz, R. (1990) . Ultimatum Bargaining Behavior: a Survey and Comparison of Experimental Results. *Journal of Economic Psychology*, 11, 417—449.

Hamilton, W. D. (1964) . The Genetical Evolution of Social Behavior. *Journal of Theoretical Biology*, 37, 1—52.

Hamilton, W. D. (1970) . Selfish and Spiteful Behaviour in an Evolutionary Model. *Nature*, 228, 1218—1220.

Henrich, J. & Boyd, R. (2001) . Why People Punish Defectors: Weak Conformist Transmission Can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas. *Journal of Theoretical Biology*, 208, 79—89.

Henrich, J. , Boyd, R. , Bowles, S. , Camerer, C. , Fehr, E. , Gintis, H. & McElreath, R. (2001) . Cooperation, Reciprocity and Punishment in Fifteen Small-Scale Societies. *American Economic Review*, 91, 73—78.

Homans, G. (1961) . *Social Behavior: its Elementary Forms*. New York: Harcourt Brace.

Isaac, B. (1987) . Throwing and Human Evolution. *African Archeological Review*, 5, 3—17.

Joshi, N. V. (1987). Evolution of Cooperation by Reciprocation within Structured Demes. *Journal of Genetics*, 66, 69—84.

Keser, C. & van Winden, F. (2000). Conditional Cooperation and Voluntary Contributions to Public Goods. *Scandinavian Journal of Economics*, 102, 23—39.

Ledyard, J. O. (1995). Public Goods: a Survey of Experimental Research. In: J. H. Kagel, & A. E. Roth (Eds.), *The Handbook of Experimental Economics* (pp. 111—194). Princeton, NJ: Princeton University Press.

Manson, J. H. & Wrangham, R. W. (1991). Intergroup Aggression in Chimpanzees. *Current Anthropology*, 32, 369—390.

Maynard Smith, J. (1976). Group Selection. *Quarterly Review of Biology*, 51, 277—283.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge, UK: Cambridge University Press.

Ostrom, E. (1990). *Governing the Commons: the Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.

Ostrom, E., Walker, J. & Gardner, R. (1992). Covenants with and without a Sword: Self-Governance is Possible. *American Political Science Review*, 86, 404—417.

Plooij, F. X. (1978). Tool-Using During Chimpanzees' Bushpig Hunt. *Carnivore*, 1, 103—106.

Price, M., Cosmides, L. & Tooby, J. (2002). Punitive Sentiment as an Anti-Free Rider Psychological Device. *Evolution & Human Behavior*, 23, 203—231.

Rogers, A. R. (1990). Group Selection by Selective Emigration: the Effects of Migration and Kin Structure. *American Naturalist*, 135, 398—413.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M. & Zamir, S. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an Experimental Study. *American Economic Review*, 81, 1068—1095.

Sato, K. (1987). Distribution and the Cost of Maintaining Common Property Resources. *Journal of Experimental Social Psychology*, 23, 19—31.

Sober, E. & Wilson, D. S. (1998). *Unto others: the Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Taylor, M. (1976). *Anarchy and cooperation*. London: Wiley.

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *Quarterly Review of Biology*, 46, 35—57.

Williams, G. C. (1966). *Adaptation and Natural Selection: a Critique of Some Current Evolutionary Thought*. Princeton, NJ: Princeton University Press.

Wilson, E. O. (1975). *Sociobiology: the New Synthesis*. Cambridge, MA: Harvard University Press.

Yamagishi, T. (1986). The Provision of a Sanctioning System as a Public Good. *Journal of Personality and Social Psychology*, 51, 110—116.

Yamagishi, T. (1988a). The Provision of a Sanctioning System in the United States and Japan. *Social Psychology Quarterly*, 51, 265—271.

Yamagishi, T. (1988b). Seriousness of Social Dilemmas and the Provision of a Sanctioning System. *Social Psychology Quarterly*, 51, 32—42.

Yamagishi, T. (1992). Group Size and the Provision of a Sanctioning System in a Social Dilemma. In: W. Liebrand, D. M. Messick, & H. Wilke (Eds.), *Social dilemmas: theoretical issues and research findings* (pp. 267—287). Oxford: Pergamon.

附录：利他惩罚的神经基础*

奎缅 费斯巴赫 特雷尔 谢尔哈默 施奈德 巴克 费尔

(Dominique J.-F. de Quervain, Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, Ernst Fehr)

许多人自愿为惩罚违反社会规范的人支付成本。演化模型和实验证据表明利他惩罚是人类合作演化的一个决定因素。我们使用 $H_2{}^{15}O$ 正电子发射断层扫描技术来检验在一个经济交易中对背叛者进行利他惩罚的神经基础。受试者可以象征性或有效地对背叛者进行惩罚。象征性惩罚并不减少背叛者的经济收益，而被实施的有效惩罚会减少背叛者的经济收益。当受试者知道背叛者滥用信任并决定对其进行惩罚时，我们可以扫描受试者的大脑。同象征性惩罚相比，有效惩罚激活了背侧纹体（dorsal striatum），它已经包含在因以目标为导向的行动而产生的奖励过程中。而且，有更强背侧纹体激活能力的受试者愿意花费更多的成本对背叛者进行惩罚。我们的发现支持人们从惩罚违反规范的人中得到满足以及背侧纹体激活反映了从惩罚背叛者中得到预期满足的假说。

人类社会合作的机制和水平在动物界是罕有其匹的。人们经常可以同一大群毫无血缘关系的陌生人合作，这些人可能以后再也不会碰到。

---

* 原文题目为 The Neural Basis of Altruistic Punishment，发表于 *Science* 305(2004) 1254—1258，胡芸译。

最近的研究表明，强互惠行为（利他惩罚和利他奖励的结合）在人类合作的演化发展中占有极其重要的位置（注 1—3）。人们经常对他人的合作和遵守规范的行为进行奖励，对违反社会规范的人进行惩罚（注 4，5）。在数千年的历史中，人类社会并没有建立法律实施的现代制度——用公正的警察和公正的法官来确保对违反规范的行为（例如经济交易中的欺诈）进行惩罚。因此，社会规范不得不通过其他手段来实施，私下的制裁就是其中之一。即便在今天的西方社会，很多规范仍然是靠私下的制裁来实施的。如果制裁在带给别的个体经济利益的同时，使实施制裁的人付出了代价，那么这个制裁就是利他的。例如，对在经济交易中欺诈的人进行制裁，以使欺诈者未来的交易对象受益，因为欺诈者现在更加意识到欺诈会被惩罚。这种意识可能可以防止将来的欺诈（注 3）。

为什么人们在没有得到物质补偿的情况下还愿意惩罚违反公认规范的人呢？我们认为个体从惩罚违反规范的人中得到了满足。一些文献也证明了这个假说。首先，最新的社会偏好（注 6—8）模型所定义的效用函数包含了对违反公正和合作规范的惩罚愿望。这些模型能比自利偏好模型更好地解释实际行为，支持了人们有愿望去惩罚违反规范行为的观点。其次，最近的人类合作演化模型（注 1，2）也表明利他惩罚行为有其长远的演化根基。这表明最近的导致人类承担惩罚他人的成本的机制是演化而来的。利他惩罚并不是一种像消化食物一样的自动反应，也不是一种基于深思熟虑、有明确目的的行为，那么人们必须有惩罚的愿望。这种诱导出有动机行为的机制说明人们从这种行为中得到了满足。大多数人在发现违反规范的行为未得到惩罚时会觉得不舒服，但一旦公正得以建立他们就感到轻松和满意。许多语言中有这样的格言警句来表明这种感觉，例如，“复仇的滋味是甜美的”。

1 研究惩罚背叛者的实验

我们利用正电子发射断层扫描技术来观察采用真实货币支付的经济

实验，以此来检验人们从惩罚背叛规范中获得满足的假说。我们的假说预测利他惩罚和与奖励过程的大脑区域的活跃程度相关。非人类灵长类（注 9—11）单一神经元的记录和人类使用货币作为奖励媒介（注 12—16）的神经成像研究可靠地表明，纹体是与奖励相关的神经回路的一个关键部分。再者，如果利他惩罚的发生是因为惩罚者预期从惩罚中得到满足，我们就应该观察到与奖励相关的脑区明显兴奋，而这个脑区又和以目标为导向的行为相关。非人类灵长类（注 17—19）的单一神经元记录提供了强有力的证据，表明背侧纹体在目标导向机制中是整合激励信息和行为信息的关键。最近的神经成像研究也支持背侧纹体与奖励决策过程相关的观点（注 20）。

在我们的实验中，两位参与者，A 和 B，以匿名的方式进行交往（注 21）。两个参与者都知道自己和一个参与者进行交往，开始时每个人都得到 10 单位的货币。如果 A 信任 B 而且 B 也以值得信任的方式行动，他们就可以充分扩大自己的收益。更具体地说，A 做第一个决策，他可以把自己的全部初始财富都交给 B（情况 1）或者他自己留着（情况 2）。如果 A 充分信任 B（情况 1），实验者就把 A 交给 B 的数量扩大到四倍，那么 B 可以得到 40 单位。这样，B 一共有 50 单位的货币，他自己的 10 单位加上被赠予的 40 个单位，此时 A 没有留下什么。接着 B 要决定是还 50 单位的一半给 A，还是什么都不还。如果 B 是值得信任的，他会送还 25 单位给 A，这样每个人都获得 25 单位货币；如果 B 留下所有的钱，他就保有 50 单位而信任他的 A 分文不得。情况 2 说明 A 不信任 B，大家都保留初始的 10 个单位（注 22）。

我们认为，如果 A 信任 B，合作和公平的规范要求 B 送还一半的收益；而如果 B 不守信留下了全部的钱，A 就视之为一种背叛规范的行为，我们认为这产生了惩罚 B 的愿望。因此，A 被赋予选择的权利，处以 B 最高 20 个惩罚点的惩罚（注 23）。在 A 获知 B 的行为后，有一分钟的时间来思考和决定是否要惩罚 B，如果要惩罚 B 就必须决定惩罚的点数。实验者在一分钟后要求 A 作出决定。因为我们对惩罚的神

经生理基础感兴趣，所以在这一分钟内对A的大脑进行了扫描。A总共和七个不同的B配对，他重复了七次上面的实验。由于A信任B而B也值得信赖的情况会带给双方相当多的收益，所以A有极大的激励去信任B；事实上，只有一个受试者在所有的七个实验中都信任B。A在七个实验的三个实验中遇到了可信赖的对手，但剩下的四个B在实验中保留所有的钱。因为我们对利他惩罚的成像感兴趣，同时为了尽可能减小辐射，我们扫描了那些B保留所有钱的实验，A只在这些实验中有惩罚B的愿望。在每个实验之间的10分钟间隔中，A填写了调查问卷，在问卷中A按照七点Likert尺度评估了B在前面实验中行为的公平性，以及惩罚B的愿望。15位健康的右撇子男士作为受试者A参与了我们的实验。由于我们关心的是A对滥用信任的反应，所以分析了信任B的14个受试者。

2　不同条件下预期的大脑活跃程度

在B保留全部收益的四个实验中，A面临着四种不同的条件。在惩罚阶段，这些条件产生的对比对衡量与奖励相关脑区的活跃程度是必要的。在称为IC（有意而且代价高）的情况中，B自己决定保留全部或者送还，这样如果B保留所有的收益，那么他就是有意滥用A的信任。而且这时惩罚对于A和B都是有代价的。每一个对B的惩罚点将花费A一个单位货币并同时减少B两个货币单位的支付。在称为IF（有意但无代价）的情况下，B同样自己决定转让与否，但是惩罚B对于A来说是没有代价的，施加于B上的惩罚并不花费A什么，而B依然在接受每个惩罚点时减少两单位的货币支付。第三种称为IS（有意但是象征性的），B还是自己决策，但是惩罚仅仅是象征意义上的，每个惩罚点都不减少A或者B的支付，这时A并不能减少B最后的收益。最后一种情况是NC（无意却有代价），B的决策是随机决定的，

这时 B 不需要为他的决策负责，但是惩罚同时对 A 和 B 有代价，每个惩罚点减少 B 两单位货币支付、A 一单位的货币支付（注 23）。为了控制顺序效应（sequence effect），这四种情况出现的顺序是随机的。

这些条件使我们可以通过计算相关条件下大脑活跃程度的差异来检验我们的假说。特别地，我们预计当 A 的信任被滥用以后，IF-IS 对比激活了与奖励相关的脑区。我们推测，A 在 IF 和 IS 条件下都有惩罚 B 的愿望，因为 B 是故意在滥用 A 的信任，不过 IS 条件下 A 不能对 B 进行实质性的惩罚。纯粹象征意义的惩罚并不令人很满意，因为惩罚背叛者的愿望无法有效地实现；在这种情况下，我们估计它的发生次数会比 IF 条件下少。

从有效惩罚中得到的满足可能有不同的心理根源。不实施惩罚的受试者可能感觉糟糕，因为背叛者逍遥法外而且所得比他们自己支付的高得多；在这种情况下，有效的惩罚避免了不断加强的负面结果。另一方面，有效的惩罚可能被视为正义的行为，受试者会因此感觉良好；在这种情况下，惩罚会不断正面强化结果。

除了有效惩罚机会的差异外，在所有情况中其他条件都保持不变，IF-IS 对比是一个检验有效惩罚的满意程度的理想方法。如果惩罚在 IF 中是令人满意的，我们预期受试者同样也愿意承担惩罚背叛者引致的成本。实际上，那些在 IF 条件下显示出与奖励相关的脑区高度活跃的受试者也应该是 IC 条件下愿意为惩罚承担最高成本的个体。此外，如果受试者合理地权衡惩罚的代价和满足的程度，也就是说，当惩罚的边际成本小于惩罚的边际“收益”的时候他才惩罚，这时 IC 条件下的惩罚也应该被认为是令人满意的。这样我们认为 IC-IS 条件下与奖励相关的区域也是活跃的。

如果 B 在 NC 条件下保留所有的钱，他并不需要为这个行为负责，因为是外加的随机机制迫使他如此。因此我们预计 A 并不认为 B 这种独占是不公平的，也没有愿望，或者只有很微弱的愿望来惩罚 B。如果没有惩罚的愿望，那么惩罚也不可能令人满意。因为这个原因，

我们预测在 IF-NC 和 IC-NC 对比中与奖励相关的区域会激活。最后，我们也能计算复合对比（IC+IF）—（IC+NC）。我们应该能观察到这一比对中与奖励相关的区域会被激活，因为在 IC 和 IF 中都有愿望和可能进行惩罚，而在 IS 中没有机会，在 NC 中没有意愿。如果对有效惩罚的意愿或者可能都没有的话，惩罚就不带来或者只带来很少的满足感。

调查问卷和实验观察支持了这些假说（图 1 的 A 到 C）。A 把 B

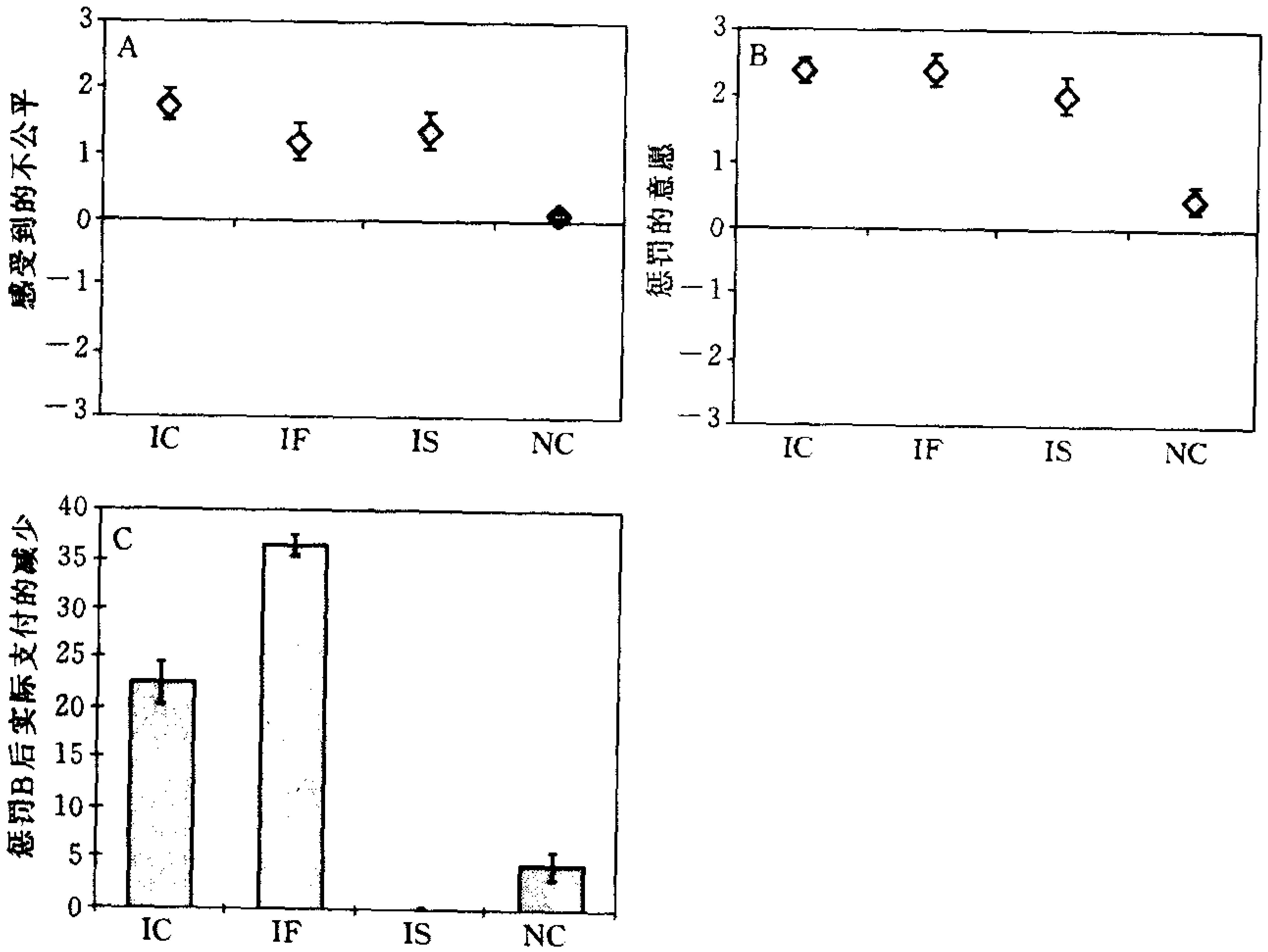

图 1　参与者 A 对 B 的感觉和对 B 实施惩罚所减少的实际支付

(A) 如果 B 保留了所有的钱，那么 A 认为这是不公平的。在 PET 扫描的 10 分钟的间隔中，参与者 A 通过在一个七点的 Likert 尺度（从 −3 到 +3）来表明他把 B 在前面实验中的行为看成是公平的还是不公平的。最大的公平用 −3 来表示，最大的不公平用 +3 来表示。这个图指出受试者之间的平均预期。(B) 如果 B 保留所有的钱，参与者 A 惩罚 B 的愿望。在 PET 扫描的十分钟间隔中，A 使用七点 Likert 尺度（从 −3 到 +3）来表明他奖励或惩罚 B 的愿望的强度。该图表明惩罚和奖励的平均愿望。(C) 如果B 保留所有的钱，B 所遭受的实际支付减少。该图表明 A 惩罚 B 后的实际支付的减少。在 IS 条件下，B 的经济支付不会减少。

在三种有意的条件（IC，IF和IS）下保留所有钱的行为看做是很不公平的举动，而在NC条件下把这个行为看成是中性的（图1A；等中位数显著性检验，$P < 0.002$ 时NC和每个有意情况的两两对比）。相似地，A在三种有意的情况下显示出强烈的愿望惩罚B，而NC下几乎没有这个愿望（图1B；等中位数显著性检验，$P < 0.012$ 时NC和每个有意情况的两两对比）。此外，A对B有意滥用他的信任的惩罚是很高的，但在NC条件下对B几乎不进行惩罚（图1C；$P \leqslant 0.001$ 时比较IC和NC的显著性检验以及比较IF和NC的显著性检验）。14个受试者中有12个对IC条件下B的独占进行了惩罚，在IF条件下全体受试者都对B实施了惩罚。这与在NC条件下形成鲜明的对比。在NC条件下14个受试者中的3个减小了B的支付，而且这三个受试者对B的惩罚是很微弱的。

3 惩罚是否激活与奖励相关的脑区的回路？

在上述对比中显示出更为活跃的区域是尾核（表1），它在我们预计的与奖励相关的活跃区域的五个对比中被激活。例如，在复合对比中最活跃的是（注6，22，4）尾核的头部（图2A；$P < 0.05$，通过多重比较来校正）。此外，尾核（图2B）血流峰值的有效范围分析表明了不同条件对其活跃程度所作的贡献：在IC和IF条件下我们观察到超过平均水平的活跃程度，这时受试者表现出强烈的惩罚愿望并且这个愿望可以得到满足；在IS和NC条件下，我们发现活跃水平低于平均水平，在这些条件下受试者要么不能满足惩罚的愿望，要么没有惩罚的愿望。尾核的这个活跃模式在IF-IS，IC-IS，IF-NC，IC-NC的单独对比中也存在（表1）。

在受试者有强烈惩罚愿望并且可以实现惩罚的情况下，尾核的活跃程度是很值得注意的，因为这一区域对奖励过程起到了很显著的作用。

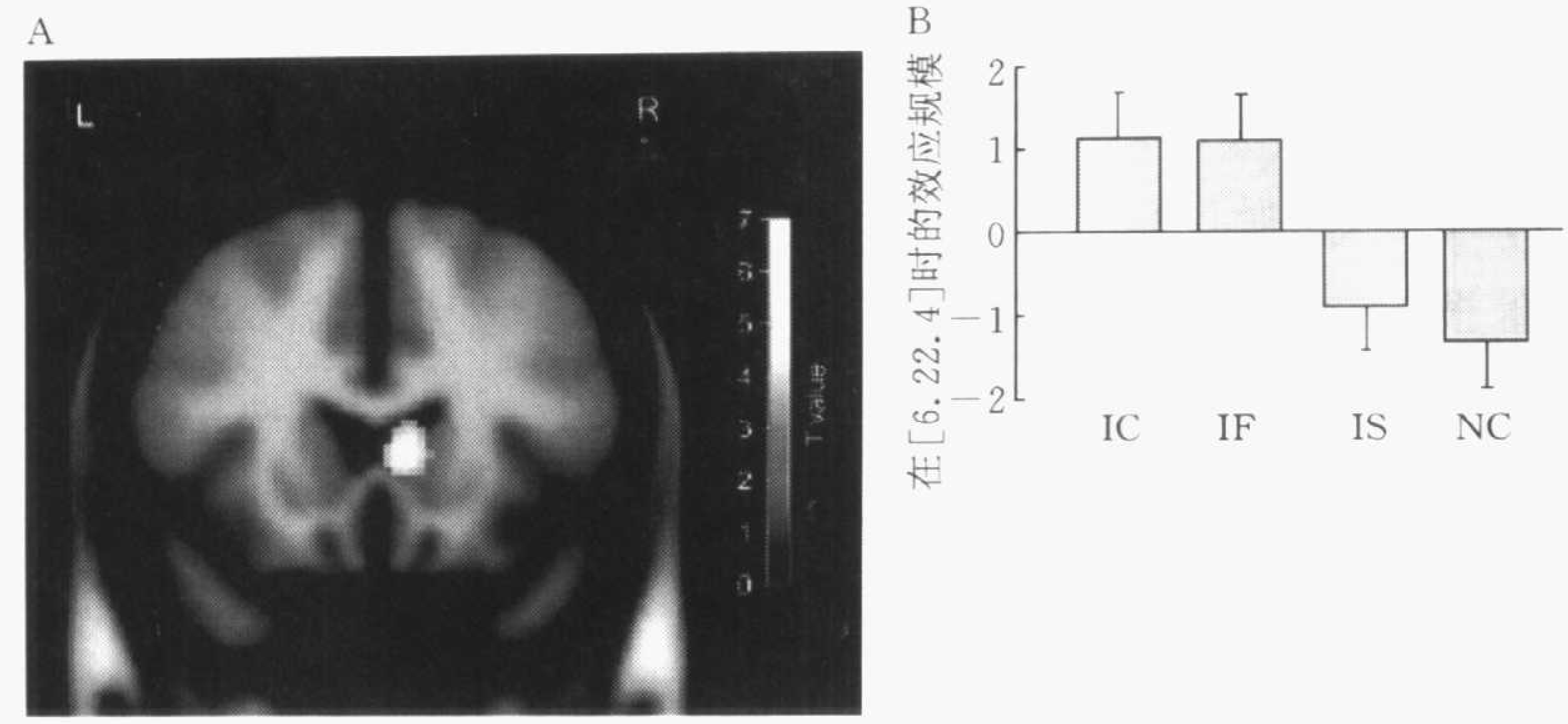

图 2

（A）受试者表明惩罚的愿望并能有效进行惩罚（IC 和 IF）相对于不能有效惩罚或没有惩罚的愿望（IS 和 NC）下尾核的活跃程度。（B）在尾核中血流量增加到最高水平的效应规模。柱状体表明相对于平均大脑激活的尾核活动。

表 1　PET 结果

对比	区域（BA）	坐标			Z 值
		x	y	z	
(IC + IF)—(IS + NC)	Caudate nucleus	6	22	4	5.11*
	Thalamus	22	−24	10	4.43*
IF-IS	Caudate nucleus	6	22	4	3.55
	Thalamus	22	−22	10	4.21
IC-IS	Caudate nucleus	6	24	2	3.70
	Thalamus	22	−22	10	4.15
IF-NC	Caudate nucleus	6	22	4	4.18
IC-NC	Caudate nucleus	6	22	4	4.23
IC-IF	Ventromedial prefrontal cortex（BA 10）	2	54	−4	4.59
	Medial orbitofrontal cortex（BA 11）	−4	52	−16	3.35

该表表明定位血流最大变化的 MNI 坐标（x，y，z）。* 表明在 $P < 0.05$ 时的显著活跃，通过多重比较纠正。否则，假设的脑区上限是 $P < 0.001$，未校正。对所有在 $P < 0.001$ 时的激活，见（注 21）。BA 是指 Brodmann 区域。IC 是指有意和代价的条件，IF 是有意和无代价的条件，IS 是有意和象征性的条件，NC 是无意但有成本的条件。x 坐标的负值表明脑的左边，MNI 是指蒙特利尔神经学院。

大鼠的损伤性实验（注 24）和其他灵长类动物（注 10， 19）的单一神经元记录实验发现这一脑区和奖励信息过程相关。人类尾核的活跃也见于几篇对奖励过程（注 12， 13， 15， 16， 25， 26）进行研究的神经成像报告中。此外，在诸如可卡因（注 27）和尼古丁（注 28）的强化刺激中也发现了尾核活跃。一些神经成像研究表明，货币刺激参数的增加与尾核的活跃程度正相关（注 14， 15）。

我们还发现了丘脑（表 1）在 IC 和 IF 情况下相对于象征意义惩罚的情况下的血流量增加。IC、IF 条件和 NC（没有惩罚的愿望）条件相比较，在 NC 条件下并没有发现丘脑被激活。在对货币激励过程（注 14， 16， 26）进行考察的神经成像研究中有发现丘脑被激活的报道。综合这些事实，我们的发现表明尾核在惩罚有意滥用信任的愿望得到满足的相关奖励过程中起到了显著作用，丘脑可能也对此起一定作用。

如果我们能指出有更强尾核活跃度的受试者会更强烈地进行惩罚，那么我们的结论将得到进一步支持。我们通过计算 IC 条件下不同受试者的实际货币惩罚和脑区兴奋之间的相关性来检验这个问题。我们发现尾核活跃（在坐标位置 [10， 26， −2]； $P < 0.001$）和惩罚投入（图 3A）之间存在明显的正相关。这一相关性有两种方式来解释：较高的惩罚会导致较强的满足感，这表明较强的惩罚会使尾核更加活跃；另一种是预期从惩罚背叛者中得到较强满足的受试者愿意在惩罚上投入较多。如果后者是正确的，则因果关系被颠倒了，因为较强的尾核活跃反映了从惩罚中得到更高的预期满足感，这反过来导致了对惩罚的较高投入。第二个解释在目标导向机制意义上考虑尾核在整合激励信息和行为信息中的重要作用是特别有意思的。

4 脑区激活与惩罚决策

我们的数据使我们可以区分这两种解释，关键在于验证在 IF 条件

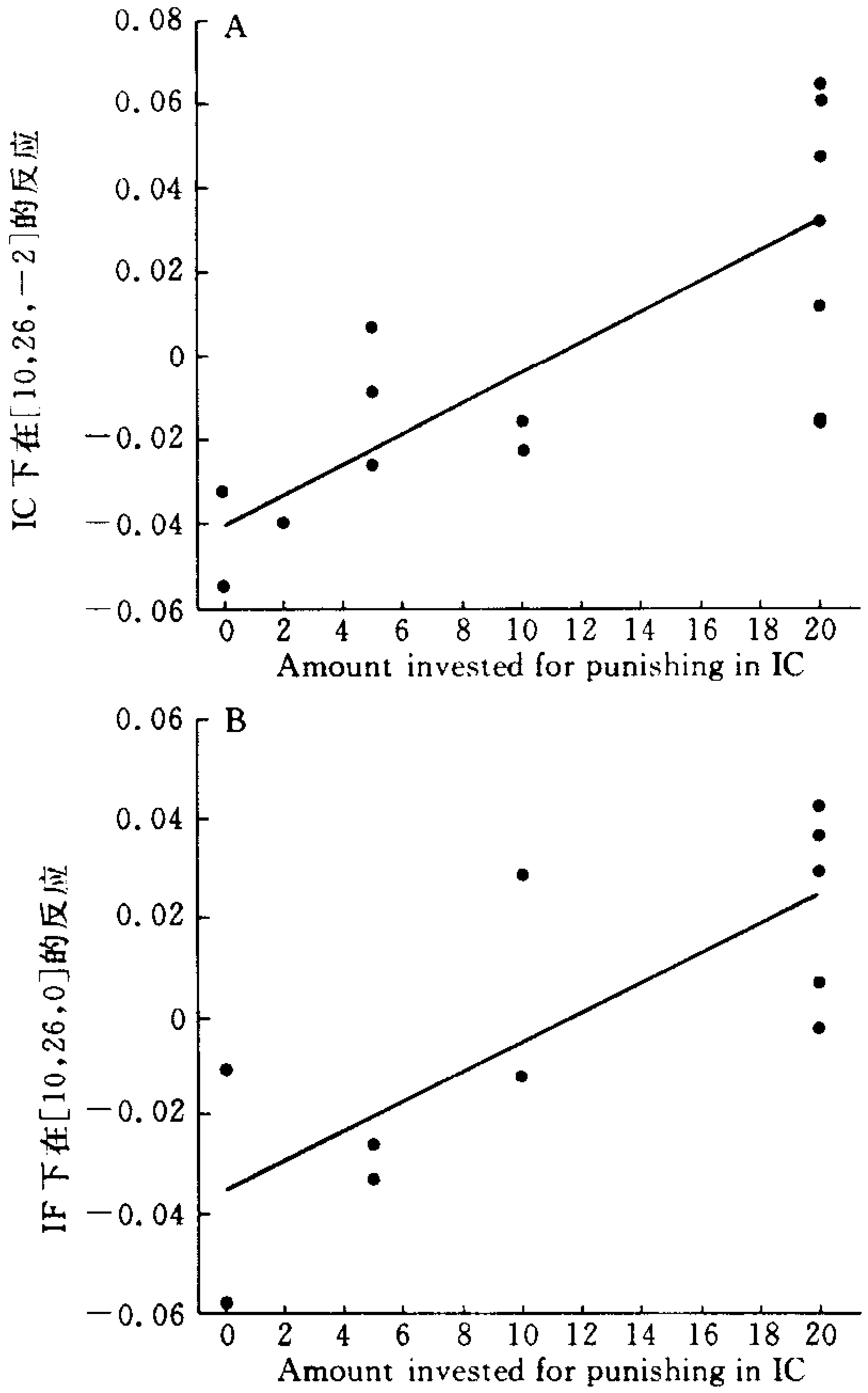

图 3

(A) 在 IC 条件下，坐标 [10, 26, −2] 上的尾核激活和实施惩罚所花的钱之间存在正相关。在 IC 条件下，有较高尾核激活的受试者会支付更多的惩罚成本。(B) IF 条件下，在坐标 [10, 26, 0] 上的尾部激活的受试者最大惩罚和在 IC 条件下受试者花费的钱之间存在正相关。有较高尾核激活的受试者在 IF 条件下比在 IC 条件下愿意为相同（最大）水平的惩罚支付更高的成本。

下实施最大化惩罚的 11 个受试者的尾核活跃情况。因为这些受试者对 B 施加了同样的惩罚，所以他们尾核活跃的差异不可能是因为惩罚程度的差异引起的。然而，如果尾核的活跃反映了对给定的惩罚水平的预

期满足感，那么不同样本之间尾核活跃程度的差异就反映了给定惩罚水平的预期满足感的差异。如果这一解释是正确的，我们就应该观察到IF中显示较高尾核活跃程度的受试者，也就是说，如果惩罚需要成本的话，那些预期从相同水平惩罚中得到更高满足感的受试者愿意对惩罚投入更多的钱。换句话说，这一解释预计在IF中实施最大化惩罚的受试者之中，那些尾核较活跃的受试者在IC条件下会花费更高的惩罚成本。IF条件下尾核的活跃程度和IC条件（图3B；$P < 0.002$）下惩罚的投入规模正相关支持了这一预测。这一发现证实了观察到的背侧纹体活跃反映了惩罚的预期满意度，这与背侧纹体是与目标导向的奖励行为有关的一个关键区域观点一致。

如果对有意背叛者的惩罚是有偿的，则参与者A在IC条件下而不是IF条件下面临的一个权衡，因为前者的惩罚是有代价的。A必须权衡惩罚导致的情感满足和惩罚所招致的货币成本，这个权衡需在追求行为目标的过程中整合几个独立的认知过程。很多证据表明，前额叶和前额脑区底部皮层参与了整合独立的认知和决策过程（注29—32）。我们的行为数据表明，IC条件下受试者面临了一个决策问题是因为在IF条件下大多数受试者都实施最大化惩罚，而IC下惩罚者的成本使惩罚显著下降（图1C；显著性检验，$P = 0.039$）。因此，我们预期在IC-IF对比中发现前额叶和前额脑区底部的兴奋。数据显示（表1和图4）前额叶腹部正中（BA 10）和前额脑区底部中部皮层（BA 11）在这一比对中被激活。BA 10的激活是有意思的，因为这个区域与在追求更高行为目标中（注33）两个或两个以上单独认知功能的整合有关。前额脑区底部中部皮层的活跃也是有意思的，因为它与屡次提到的需要对激励值编码（注34，35）的困难选择有关。这些活跃间接支持了惩罚背叛者产生满足的假说，因为否则就没有和成本相比较的收益，也不会有整合的发生。

这些结果还阐述了利他的生物学定义和心理学定义（注4）间的显著差异。根据利他的生物学定义，利他行为是施加者有代价地把利益

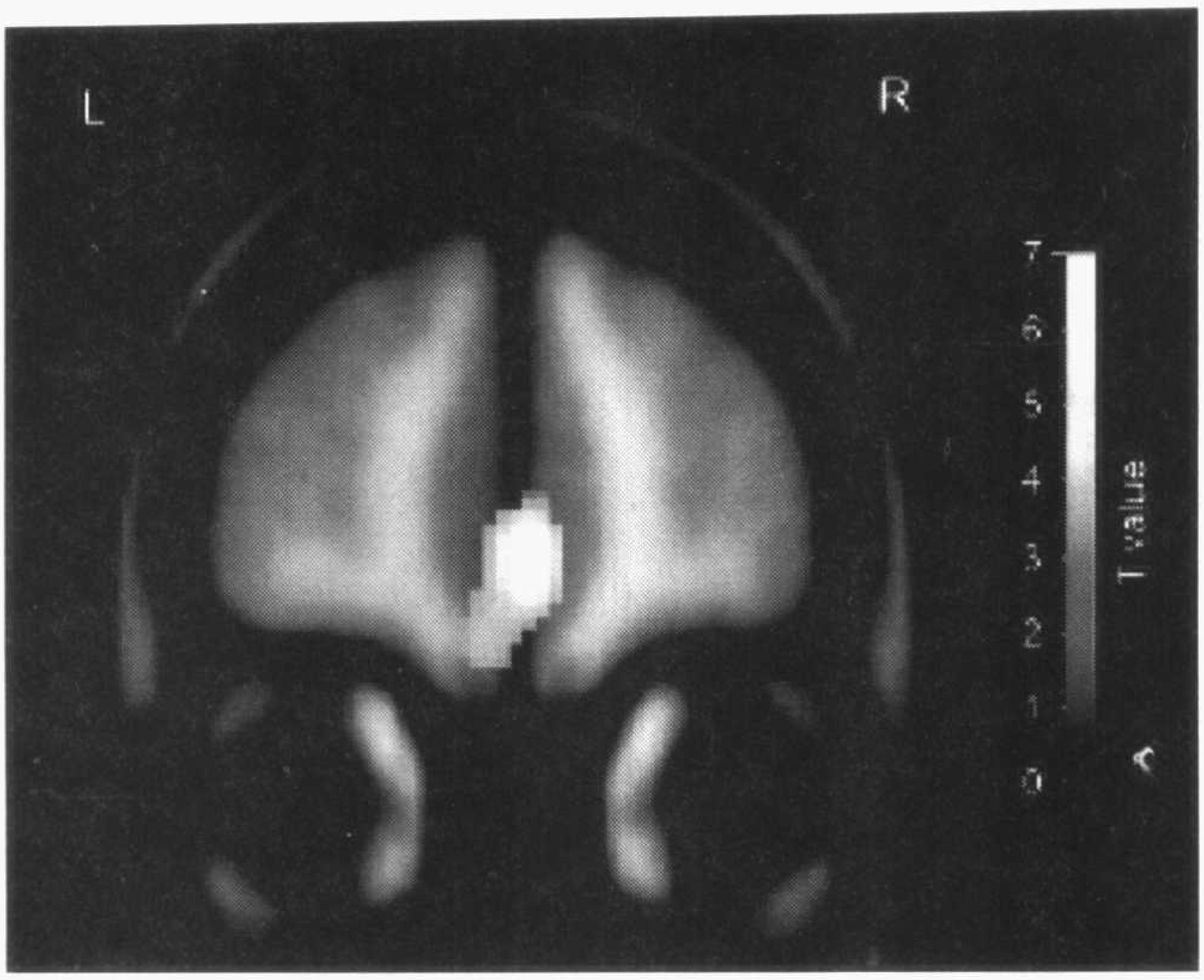

图 4　前额的皮层在整合惩罚和收益中的作用

腹正中的前额皮层和中间的圆形额在 IC 条件下相对于在 IF 条件下。

转移给其他的个体。这个利他概念的定义与这个行为是否有转移利益给他人的意愿无关，因为利他主义完全是以一系列行为的后果来定义的。这和利他的心理学定义形成对比，利他的心理学定义要求利他行为不以享受激励为基础而是受利他动机的驱使（注 36）。这样惩罚背叛者在生物学上是利他行为，因为它对于惩罚者是有代价的，而且使得被惩罚的个体以后与他人交往的时候减少背叛。尽管我们的结果表明这在心理学意义上并不是一个利他行为。

5 小　　结

我们的研究是最近尝试用“神经元经济学”和“社会行为的认知神经科学”来理解社会脑（social brain）和相关的道德情感的研究的一部分（注 37—44）。然而，这个研究是想发现对背叛者实施惩罚的神经

基础。建立适用于一大群无血缘关系的个体所组成的大群体社会规范并通过利他制裁来实施这些规范是人类区别于其他物种的一个明显特征。利他惩罚可能是解释人类社会空前水平的合作的一个关键要素（注1—3）。我们假设利他惩罚给惩罚者提供安慰或者满足，因此激活了与奖励相关的脑区。我们的设计产生了五个对比来验证假说，同时背侧纹体的前部在五个对比中都被激活，这说明尾核在利他惩罚中起着决定性作用。尾核的兴奋是非常有意思的，因为这一脑区与在预期收益激励下的决策和行动有关（注17—20）。利他惩罚中尾核的关键地位有进一步的事实支持，尾核较活跃的受试者花费了较多的惩罚成本。此外，我们的结果还揭示了这一相关性背后的原因。在惩罚成本很低的时候尾核高度活跃的受试者也愿意花费较多的资源来对背叛者实施惩罚。这样，高的尾核活跃程度就与强的惩罚意愿相关，活跃的程度反映了惩罚背叛者的预期满意程度。我们的结果支持最近发展出的社会偏好模型（注6—8），它假定人们偏好惩罚规范破坏者，阐释了利他惩罚的演化模型后面的直接机制。

注释：

[1] R. Boyd, H. Gintis, S. Bowles P. J. Richerson 2003, *Proc. Natl. Acad, Sci. U.S.A.* 100, 3531.

[2] S. Bowles, H. Gintis 2004, *Theor. Popul. Biol.* 65, 17.

[3] E. Fehr, S. Gächter 2002, *Nature* 415, 137.

[4] E. Sober, D. S. Wilson 1998, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard University Press, Cambridge, MA).

[5] E. Fehr, U. Fischbacher 2003, *Nature* 425, 785.

[6] M. Rabin 1993, *Am. Econ, Rev.* 83, 1281.

[7] E. Fehr, K. M. Schmidt 1999, *Q. J. Econ.* 114, 817.

[8] C. F. Camerer 2003, *Behavioral Game Theory*: Experiments in Strategic Interaction (Princeton University Press, Princeton, NJ).

[9] W. Schultz 2000, *Nature Rev. Neurosci*, 1, 199.

[10] P. Apicella, T. Ljungberg, E. Scarnati, W. Schultz 1991, *Exp. Brain Res.* 85, 491.

[11] O. Hikosaka, M. Sakamoto 1989, *S. Usui, J. Nwurophysiol.* 61, 814.

[12] M. R. Delgado, V. A. Stenger, J. A. Fiez 2004, *Cereb. Cortex* 14, 1022.

[13] B. Knutson, A. Westdorp. E. Kaiser, D. Hommer 2000, *Neuroimage* 12, 20.

[14] C. Martin-Soelch, J. Missimer, K. L. Leenders 2003, W. Schultz, *Eur. J. Neurosci.* 18, 680.

[15] M. R. Delgado, H. M. Locke, V. A. Stenger, J. A. Feiz, Cogit 2003, *Affect*

Behav. Neurosci. 3, 27.

[16] B. Knutson, C. M. Adams, G. W. Fong, D, Hommer 2001, *J. Neurosci*. 21, RC159.

[17] W. Schultz, R. Romo 1998, *Exp. Brain Res*. 71, 431.

[18] R. Kawagoe, Y. Takikawa, O. Hikosaka 1998, *Nature Neurosci*. 1, 411.

[19] J. R. Hollerman, L. Tremblay, W. Schultz 1998, *J. Neurophysiol*. 80, 947.

[20] J. O'Doherty et al. 2004, *Science* 304, 452.

[21] 科学在线（Science Online）提供作为支持的材料和方法。

[22] 为了和例1保持对称，当A不信任B时，参与者B仍可给予A一半的钱。然而，因为所有作为A的参与者（除了一个）都信任B，这个巧合从来没有发生过。

[23] 在所有情况下，当参与者B作了决定后，两个参与者都收到了20点额外的赠予。这个赠予允许参与者A补偿他在这些情况中惩罚他人给他耗费的成本。如果参与者A不进行惩罚，两个参与者都保留20点额外的赠予。如果惩罚对A来说不要成本，那么不论指定的惩罚点有多高，A会保留20点额外的赠予。

[24] J. A. Salinas, N. M. White 1998, *Behav. Neurosci*. 112, 812.

[25] M. J. Koepp et al. 1998, *Nature* 393, 226.

[26] M. R. Delgado, L. E. Nystorm, C. Fissell, D. C. Noll, J. A. Fiez 2000, *J. Neurophysiol*. 84, 3072.

[27] H. C. Breiter et al. 1997, *Neuron* 19, 591.

[28] E. A. Stein et al. 1998, *Am. J. Psychiatr*. 155, 1009.

[29] E. K. Miller, J. D. Cohen 2001, *Annu. Rev. Neurosci*. 24, 167.

[30] A. D. Wagner, A. Maril, R. A. Bjork, D. L. Schacter 2001, *Neuroimage* 14, 1337.

[31] D. C. Krawczyk 2002, *Neurosci. Biobehav. Rev*. 26, 631.

[32] A. Bechara, H. Damasio, A. R. Damasio 2000, *Cereb. Cotex* 10, 295.

[33] N. Ramnani, A. M. Owen 2004, *Nature Rev. Neurosci*. 5, 184.

[34] R. Elliott, J. L. Newman, O. A. Longe, J. F. Deakin 2003, *J. Neurosci*. 23, 303.

[35] F. S. Arana et al. 2003, *J. Neurosci*. 23, 9632.

[36] C. D. Batson, J. Fultz, A. Schoenrade, A. Paduano 1987, *J. Pers. Soc. Psychol*. 53, 594.

[37] R. Adolphs 2001, *Curr. Opin. Neurobiol*. 11, 231.

[38] J. D. Greene, R. B. Sommerville, L. E. Nystorm, J. M. Darley, J. D. Cohen 2001, *Science* 293, 2105.

[39] K. McCabe, D. Houser, L. Ryan, V. Smith, T. Trouard 2001, *Proc. Natl. Acad. Sci. U. S. A*. 98, 11832.

[40] J. K. Rilling et al. 2002, *J. Neuron*. 35, 395.

[41] T. Singer et al. 2004, *Neuron* 41, 653.

[42] J. Moll et al. 2002, *J. Neurosci*. 22, 2730.

[43] A. G. Santely, J. K. Rilling, J. A. Aronson, L. E. Nystorm, J. D. Cohen 1775, *Science* 300, (2003).

[44] R. Adolphs 2003, *Nature Rev. Neurosci* 4, 165.

[45] 我们非常感谢苏黎世大学、瑞士国家科学基金和MacArthur基金的经济环境、个人偏好演化和社会规范网络的支持。我们还要感谢R. Adolphs, T. Singer和L. Jäncke对我们初稿的评论，这给了我们很大的帮助。

译名对照表*

Aggressive 主动行为者

assurance game 信任博弈

bourgeois strategy 中庸策略

canonical 规范**

caudate nucleus 尾核

conformist 宿命论者（顺应者）***

conformist transmission 顺应传递

Contester 竞争者

Cooperator 合作者

costly signal 贵价信号

direct effect 直接影响

differential replication 差异复制

dorsal striatum 背侧纹体

* 该表是在《走向统一的社会科学》和《人类的趋社会性及其研究》两书编辑过程中整理而成，难免挂一漏万，望读者谅解。——编者注

** 在本书中，norm 一词也被译作规范，通常是和内化等词搭配。——编者注

*** 括号里是书中可能出现的另一种译法。——编者注

earning 所得
environmental inheritance 环境遗传
evolutionarily stable strategy 演化稳定策略
exaption 附属适应

fighting strength 战斗力
fitness 适存度（适应性）
folk theorem 无名氏定理
foraging 搜食

genetic inheritance 基因遗传
genotype 基因型
genetic correlation 遗传相关性
genetic transmission 遗传传递
genetically encoded 基因编码
globally stable 全局稳定
group selection 群体（族群）选择

heritablity 遗传可能性
horizontal transmission 水平传递
hyperfair 超公正

immediate return 立即回报
intergenerational mobility 代际流动性
inter-group 群间
interaction 互动（交往，交互）
internalization of norms 规范的内化
internalized norms 内在规范
interplay 互嬉

kin 亲缘

market integration 市场一体化（市场整合）
moonlighting game 偷袭（者）博弈

niche construction 生态位构架
norm 规范
normalized influence 正规影响
noise-to-signal ratio 噪音信号比

oblique transmission 倾斜传递
offer 出价
orbitfrontal cortex 前额脑区底部（眶额皮质）
other-regarding 他涉

parochialism 狭隘主义（地方观念）
Passive 被动行为者
payoff 支付（收益）
payoff-monotonic 支付单调
PET 正电子发射断层扫描技术
public goods 公共品
phenotype 显型（表现型）
predisposition 倾向
prefrontal 前额叶
proposer 提议者
prosociality 趋社会性 *

* 在本丛书的第一册《走向统一的社会科学》中，该名词被统译成“亲社会性”。——编者注

rate of mutation 突变率
rational actor model 理性行为者模型
rationalizability 可理性化
reciprocator 互惠者
replicator dynamics 复制动态
responder 回应者

selfish 自私
self-interest 自利
self-regarding 自涉
sequence-effect 顺序效应
socialization 社会化
stag hunt 会猎
striatum 纹体
substantial frequency 真实频率

territorial claim 领地宣告
tit-for-tat 以牙还牙
thalamus 丘脑
threat of ostracism 放逐威胁
trait 特征
trigger strategy 触发策略

Usurper 侵占者

valuation 赋值
vertical transmission 垂直传递

within group 群内

后　　记

《复杂》一书使中国学术界知道了桑塔费研究院（Santa Fe Institute）和桑塔费学派。但对于这个学术团体的后续研究工作，国内却很少进一步关注和跟追。究其原因，也许是他们大范围的跨学科研究思路，使国内学者难以适应，尤其在我国现有的科研体制和学科分野条件下。

为此，浙江大学跨学科社会科学研究中心（ICSS）在汪丁丁教授的领导下，组织了这次翻译工作。参加翻译的主要是中心的研究生和工作人员，他们是昌明、梁捷、李华芳、林水山、吴灵、周新成、毛尚熠、李欢、熊艳艳、刘征、谢家骏和胡芸，其中胡芸还承担了校对工作。这次翻译的文献近20篇，都是桑塔费学派成员的最新研究成果。考虑到篇幅问题，将分几辑出版。由于时间紧迫，加之涉及的学科领域众多，翻译上存在的问题还望读者给予谅解。

本论丛的出版得到了教育部“语言与认知研究”国家哲学社会科学创新基地和浙江大学“强所计划”的支持，特此鸣谢。此外，我们要特别感谢本书的责任编辑王志毅先生。他参照原文对本书的文字作了大量的修改和润色，并为本书配备了中英文译名对照表。

浙江大学跨学科社会科学研究中心（ICSS）

2005年4月于浙大西溪